DOCTORAL THESIS

A domain-specific framework for supporting semantic interoperability in primary and secondary use of health data on the example of the Estonian National Health Information System

Igor Bossenko

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2025

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
9/2025

A domain-specific framework for supporting semantic interoperability in primary and secondary use of health data on the example of the Estonian National Health Information System

IGOR BOSSENKO

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies
Department of Software Science

The dissertation was accepted for the defence of the degree of Doctor of Philosophy (Computer Science) on 21 January 2025

Supervisor: Dr. Gunnar Piho,
Department of Software Science,
School of Information Technologies,
Tallinn University of Technology,
Tallinn, Estonia

Co-supervisor: Prof. Dr. Peeter Ross, MD,
Department of Health Technologies,
School of Information Technologies,
Tallinn University of Technology,
Tallinn, Estonia

Opponents: Prof. Dr. Dipak Kalra,
President at The European Institute for Innovation through Health Data,
Gent, Belgium

Dr. Sulev Reisberg,
Research Fellow of Health Informatics,
University of Tartu,
Tartu, Estonia

Defence of the thesis: 11 February 2025, Tallinn

Declaration:
Hereby, I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree elsewhere.

Igor Bossenko

signature

European Union
European Regional
Development Fund

Investing
in your future

ISSN 2585-6898 (publication)
ISBN 978-9916-80-257-1 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-258-8 (PDF)
DOI https://doi.org/10.23658/taltech.9/2025
Printed by EVG Print

Bossenko, I. (2025). *A domain-specific framework for supporting semantic interoperability in primary and secondary use of health data on the example of the Estonian National Health Information System* [TalTech Press]. https://doi.org/10.23658/taltech.9/2025

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
9/2025

Tervishoiu domeenipõhise semantilise raamistiku loomine terviseandmete esmaseks ja teiseseks kasutamiseks Eesti Tervise infosüsteemi näitel

IGOR BOSSENKO

TAL TECH
KIRJASTUS

Contents

List of publications

* The present PhD thesis is based on the following publications that are referred to in the text by Roman numbers.
* The PhD thesis includes the five main publications ([I] - [V]) and five additional publications ([VI] - [X]).
* The main publications ([I] - [V]) are referenced in the section 'Publication-specific contribution' and included as appendices to the dissertation.

I Bossenko, Igor and Piho, Gunnar and Ivanova, Marina and Ross, Peeter, "TermX: The semantic interoperability, knowledge management and sharing platform," *SoftwareX*, vol. 27, 2024

II I. Bossenko, R. Randmaa, G. Piho, and P. Ross, "Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components," *Frontiers in Digital Health*, p. 30, 2024

III I. Bossenko, K. Linna, G. Piho, and P. Ross, "Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pp. 882–885, 2023

IV M. Ivanova, I. Bossenko, and G. Piho, "Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution," *Lecture Notes in Business Information Processing*, vol. 531, p. 12, 2024

V I. Bossenko, G. Piho, and P. Ross, "Modelling a patient identifier system in the Estonian National Health Information System," *Lecture Notes in Business Information Processing*, vol. 531, p. 14, 2024

VI P. Ross, J. Metsallik, K. J. I. Kankainen, I. Bossenko, C. Mäe, and M. Maasik, "Health Sense: development of a universal data model and a standard for continuity of treatment paths based on international standards of new generation health information systems," *TalTech Digikogu*, 2023. English translation: https://zenodo.org/records/14599236

VII I. Bossenko, G. Piho, and P. Ross, "Forward and Backward Compatibility Design Techniques Applying the HL7 FHIR Standard," in *HEDA@ Petri Nets*, 2022

VIII R. Randmaa, I. Bossenko, T. Klementi, G. Piho, and P. Ross, "Evaluating business metamodels for semantic interoperability with FHIR resources," in *HEDA@ Petri Nets*, 2022

IX Bossenko, Igor and Piho, Gunnar and Ross, Peeter, "TermX: A game changer in the healthcare interoperability," in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, vol. 316 of *Studies in Health Technology and Informatics*, pp. 88–89, IOS Press, 2024

X M. Marquis, I. Bossenko, and P. Ross, "RadLex and SNOMED CT Integration: A Pilot Study for Standardising Radiology Classification," *Insights into Imaging*, p. 12, 2025

Author's contributions to the publications

I In [I], I was the main author of this publication. I defined the research problem and methodology, analysed the results, prepared the figures, and wrote the manuscript.

II In [II], I was the main author of this publication. I contributed equally with the second and third authors of this publication by creating the research design and writing the introduction, methodology, discussion, and conclusion sections.

III In [III], I was the main author of this publication. I defined the research problem and methodology and conducted the literature search with the co-authors, analysed the results, prepared the figures and tables, and wrote the manuscript.

IV In [IV], I was the second author. I contributed equally with the first author of this publication by creating the research design idea, analysing the results, editing and supervising the writing of the manuscript.

V In [V], I was the main author of this publication. I defined the research problem and methodology, analysed the results, prepared the tables and figures, and wrote the manuscript.

VI In [VI], I was a member of the research group. I was responsible for modelling, FHIR profiling, terminology, and implementing FHIR in Estonia. I also analysed the results, prepared the tables, and wrote the part of the manuscript related to my area of responsibility.

VII In [VII], I was the main author of this publication. I defined the research problem and methodology, analysed the results, prepared the figures and tables, and wrote the manuscript.

VIII In [VIII], I was the second author. I contributed equally to the design of the idea of the manuscript, analysed the results, and supervised the writing of the manuscript.

IX In [IX], I was the main author of this publication. I defined the research problem and methodology, analysed the results, prepared the figures, and wrote the manuscript.

X In [X], I was the second author. I contributed equally with the first author of this publication by creating the research design. I analysed the results and supervised the writing of the manuscript.

Abbreviations

5P	Personalized, preventive, predictive, participative, precision medicine
CDA	Clinical Document Architecture
ContSys	System of concepts to support continuity of care
CTS2	Common Terminology Services
DICOM	Digital Imaging and Communications in Medicine
DHMU	Digital Health Management Unit
DS	Design Science
EHDS	European Health Data Space
EHR	Electronic health records
EMR	Electronic medical records
EIF	European Interoperability Framework
ENHIS	Estonian National Health Information System
EOSC IF	European Open Science Cloud Interoperability Framework
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability principles
FHIR	Fast Healthcare Interoperability Resources
FML	FHIR Mapping Language
HIMSS	Healthcare Information and Management Systems Society
HL7	Health Level Seven
HL7 V2	HL7 Version 2 Standard
HL7 V3	HL7 Version 3 Standard
ICD	International Classification of Diseases
IG	Implementation Guide
IHE	Integrated Healthcare Enterprise
IT	Information Technology
ISO	International Organization for Standardization
LMB	Lithuanian Medical Library
LOINC	Logical Observation Identifiers Names and Codes
OMG	Object Management Group
OMOP	Observational Medical Outcomes Partnership
OpenEHR	openEHR Standard
PCHAlliance	Personal Connected Health Alliance
ReEIF	Refined eHealth European Interoperability Framework
RIM	HL7 Reference Implementation Model
QVT	Query/View/Transformation
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
TEHIK	Estonian Health and Welfare Information Center
THO	Terminology HL7 Organization
UMLS	Unified Medical Language System
WHO	World Health Organization

Terms

eHealth	A digital technology to deliver and manage healthcare information and services.
EHR	A comprehensive digital patient-centred record that includes information from multiple healthcare providers and institutions and makes information available instantly and securely to authorised users. It provides a holistic view of a patient's health history.
FHIR	An interoperability standard that defines how healthcare information can be exchanged between different computer systems regardless of how it is stored in those systems.
Interoperability	The ability of different computer systems to exchange data with unambiguous, shared meaning.
SUSHI	FHIR Shorthand (FSH) interpreter/compiler
TermX	A knowledge development, management, and sharing platform.

Summary

The thesis is organised into the following chapters:

- Chapter 1: 'Introduction': This chapter presents the research relevance, related work in the field of semantic interoperability, focus and research questions.
- Chapter 2: 'Research Methodology': This chapter describes the applied research methodology and the developed framework used to conduct the various studies carried out as part of the thesis. It also covers the TermX development process and phases.
- Chapter 3: 'Publication-specific Contributions': This chapter provides a high-level overview of the problem, results and contributions introduced in each article. It includes the publications about the problem statement ([III], [IV], [VI], [VII], [VIII]), results ([I], [IX]) and evaluation of work ([II], [V], [X]).
- Chapter 4: 'TermX': This chapter is based on the publication [I] and provides a brief architectural overview of the artefact developed during the dissertation.
- Chapter 5: 'Discussion of Challenges': This chapter outlines the research, presents the challenges encountered in related work, summarises the results and contributions, and discusses the limitations and further research.
- Chapter 6: 'Conclusion': The final chapter summarises the thesis, discusses the importance of the interoperability domain and emphasises the role of TermX.

1 Introduction

1.1 Research relevance and background

Electronic health records (EHRs) have the potential to significantly improve the quality of healthcare outcomes, making them an essential tool for coordinated care [11]. They aim to enhance health outcomes by facilitating the delivery of healthcare services from multiple providers, ensuring that care is not delivered in silos [12]. In healthcare informatics, there is a common need for data exchange between various healthcare enterprises to better manage the quality and delivery of healthcare services [13]. However, different systems often use varied formats and standards, making seamless data sharing challenging. Interoperability has been identified as one of the greatest challenges in healthcare informatics [14]. Interoperability refers to the ability of different systems, devices, applications, or healthcare professionals to work together within and across organizational boundaries, increasing the quality and continuity of care through shared knowledge and enabling the more efficient use of that information in the healthcare process [14]. Interoperability ensures that data can be exchanged and understood across platforms.

The Refined eHealth European Interoperability Framework (ReEIF) was developed to promote and support the delivery of European public services across EU Member State borders [14]. It is positioned as an operational tool for implementers and purchasers to deploy digital healthcare information systems. The interoperability involves many aspects (Figure 1): legislation and guidelines, contracts and agreements, governance and shareable workflows, semantic and syntactic choices, applications and integration, and technical infrastructure [14].

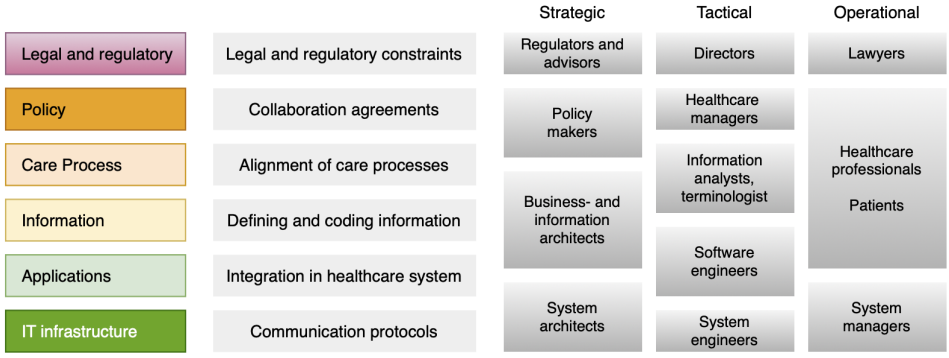

Figure 1: Refined eHealth European Interoperability Framework (ReEIF) model and stakeholders

Each country possesses unique legislation and often exhibits significant differences in clinical workflows compared to other countries. Additionally, each country typically has its own service bus for message exchange. However, the information level represents the functional description of the data model, the data elements (concepts and possible values), and the linking of these data elements to terminologies that define the interoperability of the data elements. The informational level facilitates the unification of collected health data across medical institutions and even between countries.

This thesis investigates the challenges associated with the informational layer of the ReEIF, with a particular focus on semantic and syntactic interoperability. Additionally, it explores the application layer and applications designed to facilitate these interoperability standards.

Definitions for 'syntactic and semantic interoperability' abound, but Health Level

Seven (HL7) defines them as follows: *Syntactic interoperability is the ability of one computer system to import the utterance created by another computer system and validate the utterance against a particular grammar and/or set of construction rules* [15]; *Semantic interoperability is the ability to import utterances from another computer without prior negotiation and have your decision support, data queries and business rules continue to work reliably against these utterances* [15] or the ability of different computer systems to exchange data with unambiguous, shared meaning [16]. This type of interoperability ensures that the data exchanged between systems is not only syntactically correct (i.e. correctly formatted) but also semantically meaningful (i.e. the meaning of the data is preserved and understood across systems).

The *application* level focuses on the practical implementation of interoperability solutions, ensuring that systems can work together effectively, including import, export, and data exchange of medical information [17]. Applications should be designed to seamlessly exchange data with other systems, ensuring both syntactic and semantic interoperability. Applications should be modular to allow for easy integration and scalability. Applications must ensure data security and privacy. Applications should be designed with a user-centric approach, ensuring that they are intuitive and meet the needs of the end-users [14].

Many countries have specialised organisations or departments for standardising healthcare solutions and enhancing semantic interoperability [18, 19, 20]. Their tasks include: 1) developing terminology; 2) developing logical data models to represent clinical information; 3) adapting interoperability standards such as Fast Healthcare Interoperability Resources (FHIR) [21] or HL7 Clinical Document Architecture (CDA) [22]; 4) designing data transformation between logical data models and interoperability standards; 5) creating a knowledge base and thesaurus; and 6) publishing standards and implementation guidelines.

The interoperability specialists have to support many standards at the same time, including internal data formats and/or widely accepted interoperability standards, such as HL7 V2, HL7 V3 and CDA, HL7 FHIR, DICOM, OMOP, and openEHR, as well as their various versions. Different standards approach the design of data models with varying philosophies. For example, the HL7 Reference Implementation Model (RIM) used within HL7 V3 aims to encompass the full spectrum of possible healthcare scenarios [23]. In contrast, HL7 FHIR provides a common model, but instead of constraining the scope to attempt to define a global model for all aspects of healthcare, it follows the 80/20 principle [24], designing its resources for the most common healthcare scenarios while incorporating an extension mechanism to accommodate attributes that may be absent from the models [25]. Other standards and organisations, national or global, may define data models for specific purposes. For example, OMOP defines a Common Data Model for a more limited scope of application [26].

The learning curve for eHealth interoperability is quite high due to several factors: a) complexity of standards and protocols; b) diverse systems and standards; c) regulatory and compliance requirements; and d) technical and organisational challenges [13, 27, 28].

Successful implementation of eHealth interoperability, including the development and application of data models and terminology, requires extensive domain knowledge or the use of appropriate software. User-friendly software can enable non-technical individuals to develop eHealth interoperability specifications, including terminology, data models, and transformations, more easily without needing to understand the full complexity of interoperability standards.

1.2 Overview of the proposed solution

To achieve the aim of the thesis, the author identifies the need for an interoperability platform with a rich, user-friendly interface and multilingual support. Through two research studies and ten papers, this thesis investigates several interoperability problems and presents the interoperability platform TermX—the solution for improving healthcare interoperability. As will be demonstrated by the evaluation projects conducted in publications ([II], [V], [X]), the developed platform is a valuable and reliable tool for supporting informational analysts in the process of developing and supporting interoperability solutions.

The effectiveness of the proposed platform was demonstrated through a selection of TermX by several countries as the main platform for terminology management and knowledge publishing. The TermX platform has also become an integral part of numerous master's and bachelor's theses at Tallinn University of Technology, which contribute to the enhancement of TermX. See Section 5.3 for details.

While the short-term objective of this thesis is to enhance interoperability by facilitating the coordination of digital healthcare processes through effective tools, thereby reducing the learning curve and development costs, the long-term objective is to advance the understanding of interoperability and improve the quality of electronic health records. This, in turn, aims to improve healthcare delivery and the secondary use of electronic health records.

Despite its effectiveness, the proposed solution has certain limitations (see Section 5.4 for details). These include the need for the continuous monitoring of existing electronic healthcare and modelling frameworks and competing solutions as well as the maintenance and updating of the TermX platform.

1.3 State of the art

Interoperability relies on formal standards and specifications. Organisations such as Health Level Seven International (HL7), openEHR International, and the Personal Connected Health Alliance (PCHAlliance) are creating open standards and specifications to facilitate the flow of data into electronic health records (EHRs) [11]. Achieving a consensus on system requirements and usage rules is also crucial. Integrated Healthcare Enterprise (IHE) promotes the coordinated use of established standards, such as HL7 and DICOM, to address specific clinical needs and support optimal patient care [29]. Interoperability standards can be used with different terminologies and classifiers, such as the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), the Logical Observation Identifiers Names and Codes (LOINC), the International Classification of Diseases (ICD), and others [21].

To guarantee the secure, free, and semantically correct flow of data within the EU, the ReEIF was announced in 2015 [14]. The ReEIF provides advice and guidance on improving the governance of interoperability activities, optimising processes that support end-to-end digital services, establishing inter-organisational and cross-border relationships, and ensuring that existing and new legislation does not compromise interoperability efforts. The ReEIF was designed to facilitate secure and seamless data exchange within the EU. Expected benefits include time savings, increased transparency, cost savings, improved data exchange, better data availability, better data quality, higher levels of satisfaction, improved compliance, and better decision-making [11].

Inspired by the European Interoperability Framework (basis of ReEIF), the European Open Science Cloud Interoperability Framework (EOSC IF) [30] aims to facilitate interop-

erability in the research and science domain according to FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management [31]. The EOSC IF tackles interoperability issues with the help of semantic technologies and a set of loosely coupled services and software; it provides a set of recommendations, best practices, a conceptual reference architecture, and a governance and legal structure to guide and organise the target community [32].

The Healthcare Information and Management Systems Society (HIMSS), a global advisor in health ecosystem transformation, has outlined four levels of interoperability technology: foundational, structural, semantic, and organisational [28]. The foundational level sets the requirements for connecting different systems and securely exchanging data. The structural level specifies the format, syntax, and data interpretation at the field level. The semantic level enables the use of standardised terminologies, vocabularies, and values to ensure a comprehensive understanding of the data's meaning.

The World Health Organization (WHO) SMART guidelines provide a comprehensive framework for enhancing interoperability in digital health systems [33]. The SMART guidelines are structured into layers that include the (L1) narrative, (L2) operational, (L3) machine-readable, (L4) executable, and (L5) dynamic layers, and emphasise the use of standards-based, machine-readable, adaptive, requirements-based, and testable components, which are crucial for ensuring that different digital health systems can communicate and exchange data [34]. The (L3) machine-readable layer focuses on the syntactic and semantic representation of guideline logic in digital systems. The (L4) executable layer focuses on reference applications.

The European Health Data Space (EHDS) establishes clear rules, common standards, practices, infrastructures, and a governance framework for the use of electronic health data [35]. The EHDS aims to facilitate the secure and efficient exchange and reuse of health data across the EU, benefiting patients, healthcare providers, researchers, and policymakers. It includes four interoperability layers – legal, organisational, semantic, and technical.

To guarantee consistent and conformant processes and outcomes, the specifications and principles must be based on international standards [36]. There are many health standards. Table 1 presents the classification of standards relevant in the context of personalized, preventive, predictive, participative, precision (5P) medicine [36].

ISO 23903 'Interoperability and Integration Reference Architecture' is an international standard that provides a model and framework for integrating different standards and systems; it supports the consistent and formal representation of the necessary components and their relationships [37]. It aims to facilitate interoperability and integration across various levels of complexity without the need for the continuous adaptation or revision of specifications. As a result, it provides seven interoperability layers: 1) technical for lightweight interactions, 2) structural for data sharing, 3) syntactic for information sharing, 4) semantic for knowledge sharing in computer-parsable form, 5) service/organisational for knowledge sharing at the business concept/process level, 6) knowledge-based for knowledge sharing at the domain level, 7) skill-based for knowledge sharing in individual contexts [38].

Trends in semantic interoperability can be categorised into four areas that identify challenges and research opportunities: a) frameworks designed to address semantic interoperability issues; b) the use of ontologies to resolve interoperability challenges; c) standards for achieving interoperable electronic health records (EHRs); and d) barriers and the heterogeneous nature of EHR semantic interoperability problems [39]. Several studies explore approaches to solving interoperability problems. However, there are difficulties in adopting health standards and tools for adequate data representation (ontologies,

Table 1: P5 medicine-related healthcare standards

Standards Classification	**Examples**
Architecture standards	HL7 Version 2.x/3, OMG CORBA, OMG MDA, ISO 12967 Health informatics–Service architecture (HISA), ISO 7498-2:1989, Information processing systems—Open Systems Interconnection—Basic Reference Model—Part 2: Security Architecture, ISO 13407:1999 Human-centred design processes for interactive systems
Modelling standards	OMG Unified Modeling Language (UML), ISO/IEC 19505-2:2012 Unified Modeling Language (UML), CEN 15300 CEN Report: Framework for formal modelling of healthcare security policies
Terminology and ontology standards	UMLS, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), Systematized Nomenclature of Medicine Clinical Term Ontology (SCTO), ISO 25720 Genomic sequence variation markup language, ISO/IEC 2382-8:1998 Information technology—Vocabulary—Part 8: Security, CEN-ENV 13608-1:2000 Health informatics—Security for healthcare communication—Part 1: Concepts and terminology, ISO 13940:2015 Health informatics—System of concepts to support continuity of care, Logical Observation Identifiers Names and Codes (LOINC), Unified Code for Units of Measure (UCUM)
Communication standards	ISO/IEC 7498-1:1994 Information technology—Open Systems Interconnection—Basic Reference Model: The Basic Model, HL7 V2.x/3, HL7 FHIR (Fast Healthcare Interoperability Resource), X12 EDI, UN EDIFACT, H.PRIM, xDT, Odette FTP, CEN 13606 Electronic healthcare record communication, ISO/IEEE 11073 Health informatics—Point-of-care medical device communication, ISO 17113 Health informatics–Exchange of information between healthcare information systems–Development of messages, CDISC and DICOM specifications, Classification Markup Language (ClaML), EN ISO 27269:2022 Health informatics-International patient summary (ISO 27269:2021)
Policy, security, and privacy standards	ISO/IEC 2700 Information security management, ISO 22600:2014 Health informatics–Privilege management and access control, ISO 17090 Public key infrastructure, ETSI TS 101733 Electronic Signature Formats, ASTM E1987-98 Standard guide for individual rights regarding health information, CEN 13608 Security for healthcare communication, CEN 13729 Secure user identification-Strong authentication using microprocessor cards, ISO 25237:2017 Health informatics—Pseudonymization, ISO/IEC PDTS Pseudonymisation Practices for the Protection of Personal Health Information and Health Related Services, ISO/IEC 27018:2019 Information technology—Security techniques—Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors, ISO/IEC 29151:2017 Information technology—Security techniques—Code of practice for personally identifiable information protection, ISO 21298:2017 Health informatics–Functional and structural roles, ISO/IEC 9594-8:2008, Information technology—Open Systems Interconnection—The Directory: Public-key and attribute certificate frameworks, ISO/IEC 9798-3:1998, Information technology—Security techniques—Entity authentication—Part 3: Mechanisms using digital signature techniques, ISO/IEC 10181-1:1996, Information technology—Open Systems Interconnection—Security frameworks for open systems: Overview, ISO/TS 17090-1:2013 Health informatics—Public key infrastructure—Part 1: Overview of digital certificate services, ENV 13729:1999, Health informatics—Secure user identification for healthcare strong authentication using microprocessor cards, ISO 21091:2013 Health informatics—Directory services for healthcare providers, subjects of care and other entities, ISO/IEC 15408-1:2009 Information technology—Security techniques—Evaluation criteria for IT security—Part 1: Introduction and general model
Safety standards	CEN 13694 CEN Report: Safety and security related software quality standards for healthcare, ISO/DTS 25238 Classification of Safety Risks, IEC 82304-1 Health Software–Part 1: General requirements for product safety. IEC 82304-2 Health Software–Part 2: Health and wellness apps–Quality and reliability
Identifier and identification standards	LOINC, ASTM E1714-00 Standard guide for properties of a Universal Healthcare Identifier
Document standards	HL7 V3/CDA (Clinical Document Architecture), DICOM SR (Structured Reporting), HL7 FHIR Bundle+Composition
Data representation (visualisation) standards	HTML, PDF, PDF/A, MS Word, ClaML
Encoding standards	XML, JSON, ASN.1, ER7, xDT
Character representation standards	ASCII, EBCDIC, Unicode

databases, clinical models) that ensure the efficient management of data by healthcare professionals [16].

There is no consensus on a global standard for electronic health records [16]. However, last year's trends in the standard choice show the selection of interoperability frameworks such as openEHR and especially HL7 FHIR [16, 40].

Another problem is the lack of a common, countrywide, well-defined interoperability framework for managing eHealth solutions [11]. Since the use-case approach is the underlying methodology for documenting user needs, a practical approach to achieving interoperability can be summarised in the following steps:

- Identify use cases from an end-user perspective
- Select profiles and standards that support the use case
- Refine data content, including master files, and terminology
- Prepare implementation guides
- Organise component interoperability and cross-implementer connectivity testing
- Educate end-users

- Support communities of practice to promote sustainable standards-based implementation

For each of the proposed use cases, rules need to apply to determine how to accomplish interoperability at each of the following levels [11]:

- Semantic: to make sure, through the development and use of standardised vocabularies and formats, that the meaning of exchanged data and information is well understood by the different parties, resolving any possible ambiguities regarding the notions in the healthcare domain.
- Technical: to ensure the use of formal technical specifications and widely accepted and used standards and tools.

The several use cases concerning digital transformation into FHIR [41, 42] describe processes, outcomes, and lessons learned from government digital transformation projects. Most of the use cases describe the process of migrating from legacy systems to FHIR. There are solutions for developing terminology and guidelines, and there are FHIR servers, but individually, they are not enough to launch or modernise a national health digitalisation program.

Numerous programs exist to perform specific tasks in the FHIR ecosystem, such as Ontoserver for terminology [43], SUSHI for logical models [44, 45], Forge for FHIR profiling [46], FHIR validator for data transformations [44], UMLS for thesaurus [47], Implementation Guide (IG) Publisher [48, 44], and Simplifier [49] for Implementation Guide and publishing.

Many of these software programs demand extensive human expertise. Frequently, these software programs operate in isolation, lacking integration within a cohesive ecosystem. Achieving interoperability among them poses a substantial challenge.

1.4 Estonian National Health Information System

The Estonian National Health Information System (ENHIS) maintains lifelong health records of all Estonian citizens and provides a comprehensive set of healthcare services, including discharge summaries, referrals, e-prescriptions, and the national appointment system. The development of ENHIS began in 2005. It became operational in 2008 and is based on HL7 V3 and CDA standards [50].

Health data is collected by healthcare providers and stored as primary data in their software systems in various formats [51]. The primary use of health data refers to the utilisation of health information for direct patient care [52]. Primary health data typically consists of medical records, diagnostic test results, health conditions, and treatment plans, which are essential for providing accurate and effective patient care.

Healthcare providers compose CDA documents and share them with the ENHIS [53]. The ENHIS acts as a central database, enabling HL7 CDA document exchange and storage. It also facilitates the secondary use of health data by aggregating data from different types of documents and sources, enabling care planning, decision-making, researching, and policymaking. Additionally, the ENHIS provides registries and portals for the collection and use of primary data [50].

The biggest drawbacks of the HL7 CDA document-based approach are the timing of information sharing and the transformation of healthcare data between provider systems and the ENHIS. Documents are generally shared once all the agreed data elements have been precisely filled in and the necessary confirmations received. This leads to data duplication and difficulties providing timely and complete data to key users and hinders the

efficient use of information by all stakeholders, including healthcare professionals. The absence of a standardised approach for continuity of care, along with the lack of a comprehensive terminology database and streamlined data models, further complicates data exchange and limits the secondary use of valuable health information [III].

Today, Estonia is transitioning the ENHIS from a document-based approach to an event-based approach by utilising the HL7 FHIR standard [III, V]. As a result, new services are emerging that are responsible for collecting, storing, and using primary data [V].

1.5 Focus and aim

Our focus is on semantic interoperability in healthcare, ensuring that various healthcare software systems can exchange and accurately interpret patient data, thereby enhancing patient care and minimising errors.

According to *ISO 18308: Requirements for an electronic health record architecture*, an EHR should be designed to integrate health records in a format that can be processed, securely stored, and communicated [54]. It should employ a widely accepted information model and terminology for data exchange, ensuring that the data is accessible to authorised users [11]. Such a system aims to provide efficient, high-quality, and secure integrated healthcare for patients.

Our main goal is to achieve semantic interoperability in EHRs for primary and secondary use cases. The additional goal of this research is to hide the complexity of semantic interoperability and the related standards and facilitate the participation of non-technical analysts in the development of terminology, data models, and transformations.

The desired outcome should be achieved by deeply understanding the syntactic and semantic aspects of the informational layer. This involves providing software programs with advanced export/import functionality and a rich user interface that reduces the time needed to learn the basics of interoperability, terminology, modelling, and transformations. Additionally, this outcome optimises human resources and simplifies adaptation to standards.

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

- Discussion on the state of the art of semantic interoperability in electronic health records exchange.
- Recognition of the main approaches, solutions and programs commonly used to achieve semantic interoperability in EHR (electronic health record) systems.
- Proposal of a solution suitable for non-technical staff to solve semantic interoperability problems, combining international standards and existing tools and developing missing tools.

This work aims to answer the following primary research question (RQ): *"Which solution enables non-technical staff to implement semantic interoperability in electronic health records in accordance with international healthcare standards?"*.

To provide a clear scope and decrease complexity, this RQ was split into three sub-research questions as follows:

- *sub-RQ1*: How can we manage terminology?
- *sub-RQ2*: How can we design informational models and transform data?
- *sub-RQ3*: How can we evaluate the solution?

Table 2: Mapping of associated research questions and publications

Research question	Publications
RQ	[I], [II], [III], [IV], [V], [VI], [VII], [IX], [VIII], [X]
sub-RQ1	[I], [IV], [VI]
sub-RQ2	[I], [II], [III], [VI], [VII], [VIII]
sub-RQ3	[II], [V], [X]

Table 2 presents the mapping of each sub-RQ to the corresponding publications that contribute to its answer.

2 Research methodology

This summary is composed based on ten papers and two research studies. I wrote six papers as the first author and three as the second author in collaboration with the master's and bachelor's students I supervised. The research study 'Health Sense' is the result of a group of researchers; the report was published in Estonian and the most valuable chapters for my thesis were translated into English (publication [VI]). The study 'Project TermX' is a research and development project under the Enterprise Estonia programme for applied research.

Publications [III], [IV],[VI], and [VII] contribute through the analysis of the state of the art in model and terminology design. Publication [VIII] shows that data transformations with existing tools require extensive programming experience and are not suitable for non-technical analysts. Publications [I] and [IX] describe the design principles and architecture of TermX—the interoperability platform developed during my doctoral study. Publications [V] and [X] evaluate the development of terminology using TermX. Publication [II] evaluates the development of informational models and data transformations using TermX.

2.1 Research process

The study aims to improve semantic interoperability in EHR systems by designing techniques and artefacts that domain experts can use to specify and validate data terminology, data models, and transformation rules with only minimal technical expertise and skill needed. The design science (DS) methodology [55, 56], widely accepted as an information systems research method, was adhered to during the study. In the design-science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the building and application of the designed artefact. The designed artefact needs to solve a specific problem that is rigorously defined, formally represented, coherent and internally consistent, and comprehensively evaluated [57].

DS is part of the engineering cycle (Figure 2) and includes the problem investigation, treatment design, and treatment validation phases. The treatment implementation phase is not part of DS but forms an engineering cycle along with the DS phases. Design is essentially a research process to discover an effective solution to a problem [57]. Problem-solving can be viewed as utilising available means to reach desired ends while satisfying laws existing in the environment. Design science is inherently iterative.

Based on the above-mentioned three pillars, I utilised the design science research process shown in Figures 2 and 3, resulting in the framework described in Section 2.2 and the TermX software described in Sections 2.3 and 4.

This thesis reports three DS cycles and therefore also three engineering cycles. In the first cycle, we designed and developed the TermX tool. In the second cycle, we evaluated the TermX terminology server by designing the code systems for the radiology classifier and patient identifier domain. In the third cycle, we evaluated the TermX tool by designing the techniques and reusable WYSIWYG components for data transformation rules and maps from CDA to FHIR in the ENHIS. Outside this study, TermX was deployed in the two production environments of the Lithuanian Medical Library (LMB) and the Estonian Health and Welfare Information Center (TEHIK).

The design-science paradigm seeks to extend the boundaries of human and organisational capabilities by creating new and innovative artefacts. While the implementation of the artefact (TermX tool) is not part of DS but part of the engineering cycle, Figure 2 includes its implementation to illustrate the place and role of the TermX tool's develop-

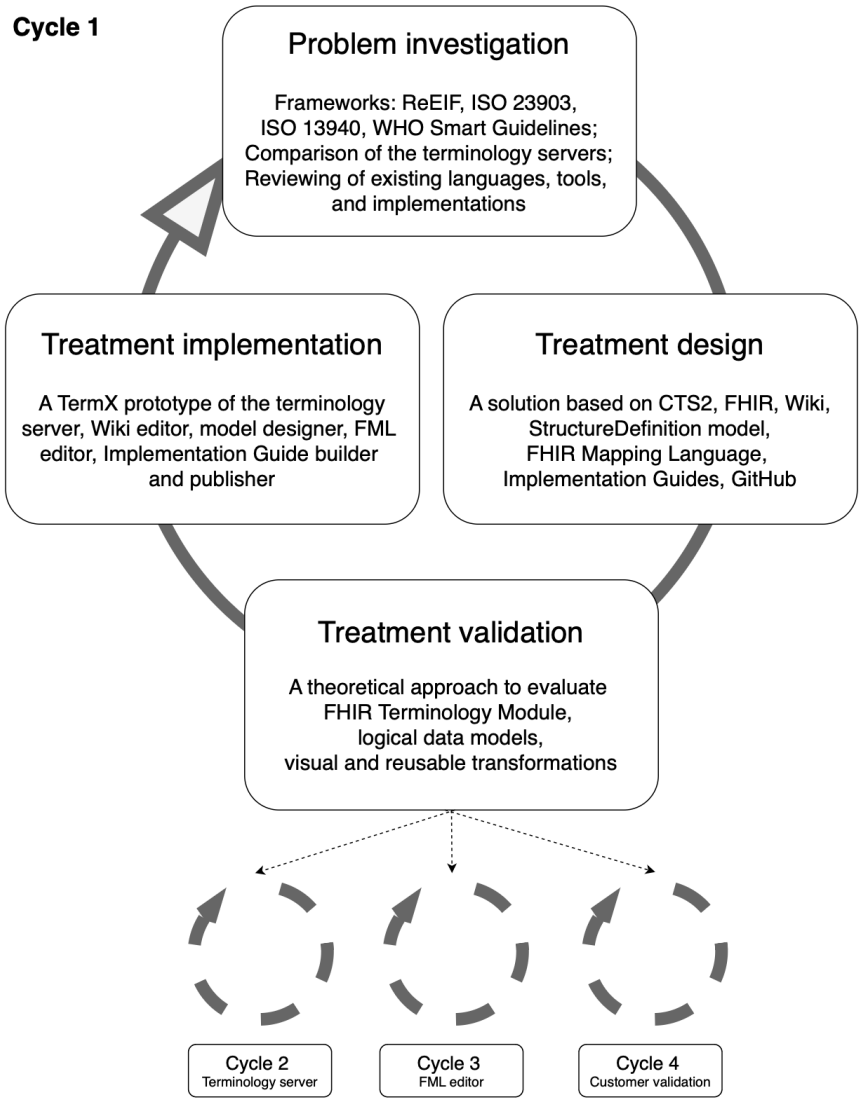

Figure 2: Research process

ment in our study. We designed TermX according to the DS methodology, encompassing the following steps: 1) investigating a problem, its relevance, and its research rigour by reviewing published papers on existing interoperability standards and frameworks, terminology servers, data transformation languages, tools, and implemented projects (see Sections 1.3 and 5.2); 2) designing the TermX tool [I] with domain experts from various countries; and 3) developing the validation criteria and planning the new DS cycles for evaluating TermX modules.

DS problems are improvement problems. The design process is a sequence of expert activities that produce an innovative product (i.e. the design artefact). The evaluation of the artefact then provides feedback information and a better understanding of the problem to improve both the quality of the product and the design process [57].

In the second cycle, the main focus is to evaluate the TermX tool by designing terminol-

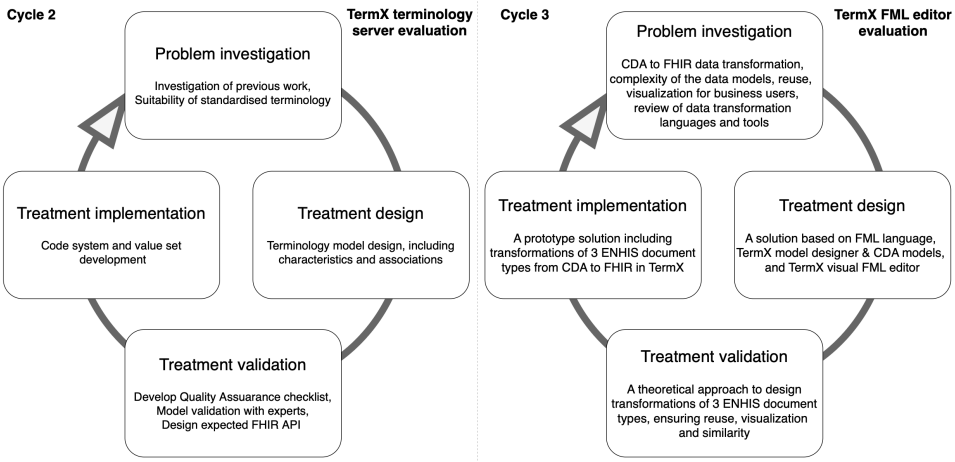

Figure 3: TermX terminology server and FML transformation editor evaluation research

ogy for real-life use cases, such as the development of the classifier for Estonian radiology procedures [X] and patient Identifier Domain [V] for the ENHIS Patient Registry.

In the third cycle, the main focus is to evaluate the TermX tool by designing visual reusable transformation components that domain experts can use for CDA to FHIR transformations [II]. We also generalise the transformation components' development process as techniques for developing reusable transformation components using TermX and explain the relevance of our research in the EHDS ecosystem, including how the proposed approach supports federated semantic interoperability.

In addition to the three cycles, TermX was deployed and validated in the customer environments of the LMB and TEHIK. This fourth cycle belongs to the engineering part and is not covered in the thesis.

Table 3: Distribution of papers by design cycles and phases

Cycle	Phase	Publications
1	Investigation	III, IV, VI, VII, VIII
1	Design	I, IX
1	Implementation	I
2	Investigation, Design	V, X
2	Validation	X
3	Investigation, Design, Evaluation	II

This thesis aims to improve the efficiency and accuracy of data exchange within healthcare systems. By leveraging DS principles, we focus on creating practical solutions that address real-world challenges faced by healthcare professionals. Our approach includes iterative cycles of design, implementation, and evaluation to ensure that the developed artefacts meet the needs of end-users and stakeholders. The ultimate goal is to enhance the quality of healthcare delivery by enabling seamless data interoperability across systems and platforms.

2.2 Framework design

Design science is inherently a problem-solving process [55]. The fundamental principle of design-science research is that knowledge and understanding of a design problem and its solution are acquired in the building and application of an artefact. That is, design-science research requires the creation of an innovative, purposeful artefact [I] for a specified problem domain [III, IV, VI, VII]. Since the artefact is purposeful, it must yield utility for the specified problem. Hence, a thorough evaluation of the artefact is crucial [II, X]. Novelty is similarly crucial since the artefact must be innovative, solving a heretofore unsolved problem or solving a known problem in a more effective or efficient manner [I, II]. In this way, design-science research is differentiated from the practice of design. The designed artefact itself needs to solve a specific problem that is rigorously defined, formally represented, coherent, and internally consistent [I]. The process by which it is created, and often the artefact itself, incorporates or enables a search process whereby a problem space is constructed and a mechanism is posed or enacted to find an effective solution [IV]. Finally, the results of the design-science research must be communicated effectively [IX], both to a technical audience (researchers who will extend them and practitioners who will implement them) and to a managerial audience (researchers who will study them in context and practitioners who will decide whether they should be implemented within their organisations).

We used the Health Sense research study [VI], evaluations of existing terminology servers [IV] and data transformation tools [VIII], and explored questions about forward/backward compatibility [VII], migrations between interoperability standards, practical implementations [III], and interviews with experts to develop the eHealth interoperability taxonomy (see Figure 4) applied to the thesis.

The developed taxonomy is based on the ReEIF 6-layer model and outlines the comprehensive scope of the thesis. At the highest level, it includes interoperability frameworks for the healthcare domain, such as the *Refined eHealth European Interoperability Framework (ReEIF)* and the *European Health Data Space (EHDS)*. These frameworks focus on establishing the legal framework, data protection, compliance, legal agreements, and accountability. While the ReEIF and EHDS are often part of regulations in European countries, the *WHO Smart Guidelines* may be part of strategic planning and policy development. Additionally, the ReEIF and EHDS can be part of the Policy Layer in non-European countries. The scope of the thesis is limited to the roles of business analysts, national standards developers and solution architects, who are responsible for implementing regulations and policies. Several countries have established a National Office of Coordination or a Digital Health Management Unit responsible for coordinating the digitisation of care processes. These units are tasked with terminology management, the development of clinical data models, the adaptation of interoperability standards, the transition from one interoperability standard to another, and the development and management of the knowledge base and standards publication. They are responsible for the selection of healthcare standards and frameworks, their integration, quality, and continuity of care. This taxonomy incorporates the standards and terminologies used in the thesis, including the HL7 standard family (V2, V3, CDA, FHIR, CTS2) and terminology (THO), the ISO 13940 standard for continuity of care, openEHR, Integrated Healthcare Enterprise (IHE), DICOM, SNOMED, LOINC, and the WHO terminology family. The *ISO 23903: Interoperability and integration reference architecture* standard provides a model and framework for integrating different standards. We apply techniques that support the standards of the information layer and utilise the most popular open-source components to implement the solution. The list of open-source solutions includes TermX—a platform for interoperability, knowledge

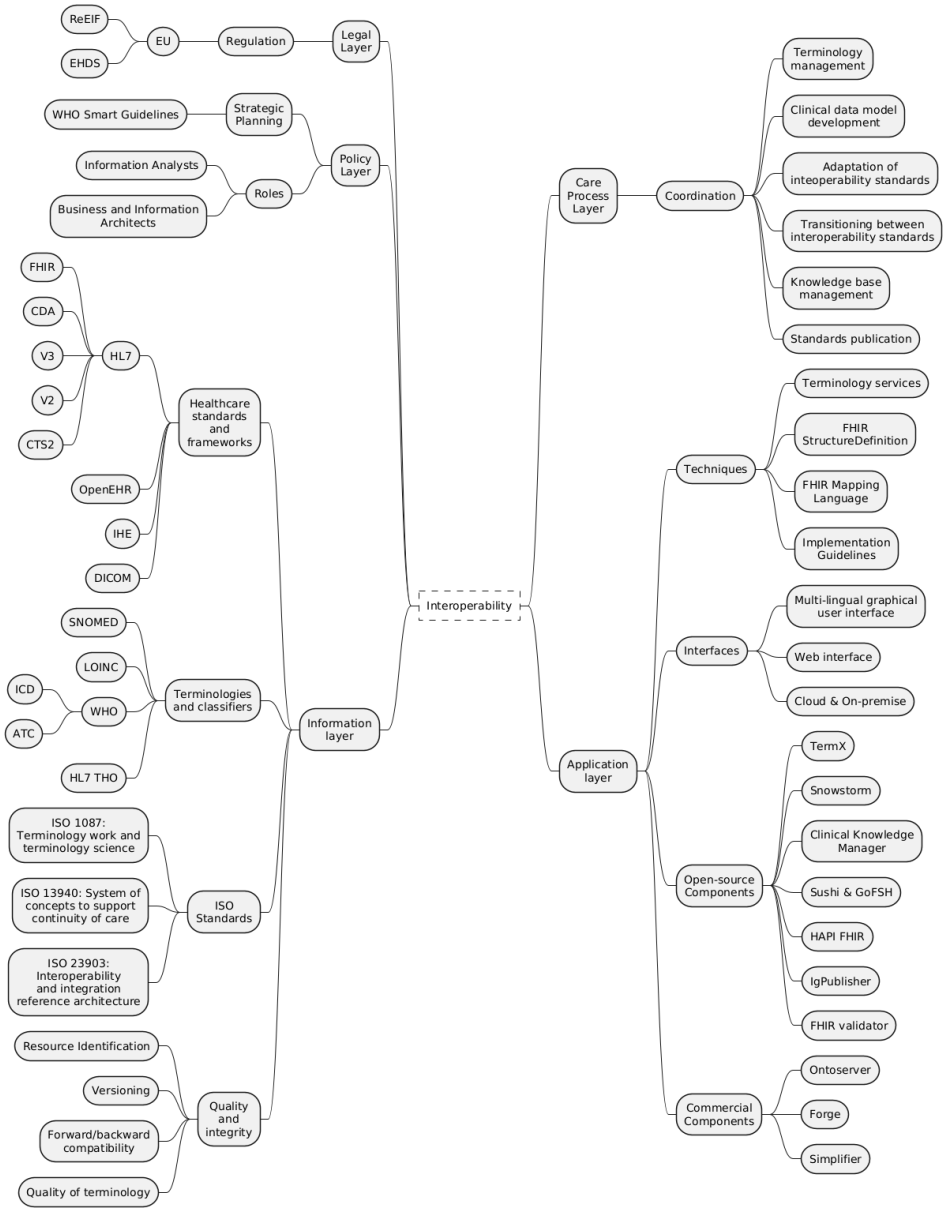

Figure 4: Proposed taxonomy for eHealth interoperability

management, and data sharing developed during this thesis. TermX is designed to simplify the complex eHealth interoperability landscape and support non-technical roles in the implementation of healthcare solutions.

Since the number of standards is very large, this thesis is limited to the latest HL7 standard – HL7 Fast Healthcare Interoperability Resource (FHIR), and related standards and terminologies. HL7 FHIR [58] is an international healthcare information exchange standard that provides a range of predefined resources, the possibility to expand these resources,

and a framework for the exchange of these resources between interested parties [VII]. A full list of standards and frameworks used in the thesis is given in Table 4.

Table 4: Thesis-related healthcare standards and frameworks

Standards classification	Used standards and healthcare frameworks
Architecture standards	HL7 FHIR, HL7 V3, OpenEHR
Modelling standards	HL7 Common Terminology Services (CTS2), HL7 Reference Information Model (RIM), FHIR Reference Model, openEHR Reference Model, OMG MOF Query/View/Transformation (QVT), System of concepts to support continuity of care (ContSys)
Terminology and ontology standards	Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), Logical Observation Identifiers Names and Codes (LOINC), International Classification of Diseases (ICD), HL7 THO (Terminology HL7 Organization)
Communication standards	HL7 FHIR
Identifier and identification standards	IHE (Integrated Health Enterprise), HL7 THO, RFC3043 - A URN Namespace for People and Organizations
Document standards	HL7 V3/CDA (Clinical Document Architecture), HL7 FHIR Bundle and Composition
Data representation (visualisation) standards	HTML, PDF, CSV, Excel, PNG
Encoding standards	JSON, XML
Character representation standards	ASCII, Unicode

2.3 TermX development

The development of TermX began during the Health Sense research project [VI]. One of the outcomes of the research was the need for a universal terminology database. The research revealed that the universal data model for terminology is specified by the HL7 CTS2 standard [59]. At the same time, the study of existing terminology servers showed that they do not use this standard and have many other limitations [IV]. To test the functionality of CTS2, a prototype was created that can create code systems and value sets through an API and user interface [60]. The prototype has demonstrated that the CTS2 standard fulfils the requirements for a universal terminology database as outlined in the publication [61]. This includes the formal definition of concepts, the versioning of concepts, code systems, and value lists, and support for the FHIR API.

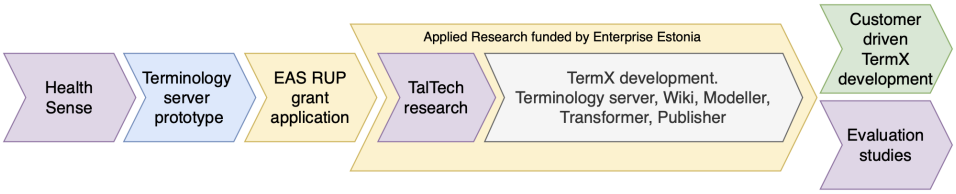

Figure 5: Key milestones in TermX development

Since the prototype development was successful, it was decided to continue developing the platform, which was named TermX. A request for a grant from the Programme for Applied Research by Enterprise Estonia was submitted and approved. The TermX development project was conducted in cooperation between TalTech and a commercial company. The project began with a TalTech investigation into the modelling, transformation, publishing, and clinical aspects of terminology, interviews with nine world-leading experts, and an analysis of existing tools. The evaluation studies (as separate DS cycles) were initiated to validate the TermX functionality of the terminology server, as well as the modelling and transformation tools, after reaching Technology Readiness Level 5 (by the European Commission Technology Readiness Level classification). At the same time, TermX participated

in two national competitions and was selected as the platform for national terminology development in Lithuania by the LMB and in Estonia by TEHIK. As a result, many features were added to TermX within these projects.

The TermX development phases are demonstrated in Figure 5.

3 Publication-specific contributions

This section presents the semantic interoperability topics introduced in each publication to answer the research questions. The aim is to give a high-level overview of the problems, the novelty of the solution, and the contribution to the thesis. The order of the publications is not chronological according to publication date but represents the recommended reading order.

3.1 Comparative analysis of clinical terminology servers: a quest for an improved solution [IV]

The paper provides a comprehensive evaluation of existing clinical terminology servers and proposes the need for an enhanced solution to meet the evolving demands of modern healthcare.

The study emphasises the critical role of clinical terminology servers in ensuring seamless communication and interoperability across healthcare systems. It evaluates the strengths and weaknesses of prominent terminology servers, such as Ontoserver, Snowray, Rhapsody, TermSpace, Snowstorm, Hermes, and Apelon DTS, against a set of predefined criteria. These criteria include support for standardised terminology classifications, CRUD operations, terminology import capabilities, HL7 FHIR integration, multilingual support, versioning mechanisms, web interface usability, and licence type. The findings reveal significant gaps and limitations in the current solutions, demonstrating that there is a need for a new, improved terminology server.

3.1.1 Problems addressed

The current clinical terminology servers face significant challenges in supporting standardised terminologies, offering multilingual capabilities, and integrating with modern interoperability standards such as HL7 FHIR. These limitations create barriers to effective communication and data exchange in healthcare, making the development of a more robust and adaptable solution essential. To address these challenges, the study emphasises the need for efficient terminology management, including CRUD operations (create, read, update, delete), versioning, and import/export options, which are vital for maintaining clinical terminologies. It also notes the limitations of existing servers in importing and exporting terminologies in various formats, a crucial requirement for integrating external data sources. Additionally, the study stresses the importance of compliance with modern standards such as HL7 CTS2 and HL7 FHIR, revealing the shortcomings of current servers in adhering to these protocols. Finally, it examines the usability of existing server interfaces, calling for a more user-friendly, multilingual design to enhance interaction and navigation for healthcare professionals.

3.1.2 Novelty

The study introduces a structured approach to evaluating clinical terminology servers based on essential criteria, such as standardised terminology support, HL7 standards support, CRUD operations, and multilingual capabilities. It highlights the need for a new terminology server that addresses the identified gaps, offering enhanced import capabilities, adherence to the latest standards, and a user-friendly web interface.

Table 5: Terminology server compliance with criteria

Criteria	Ontoserver	Snowray	Rhapsody	TermSpace	Snowstorm	Hermes	Apelon DTS
Standardised terminology classification	SNOMED LOINC ICD	SNOMED LOINC ICD	SNOMED LOINC ICD	SNOMED	SNOMED LOINC ICD	SNOMED	SNOMED LOINC ICD
CRUD operations	+	+	+	+	N/A	+	N/A
Code System import	+	+	N/A	N/A	+	N/A	N/A
Supported FHIR release	N/A	N/A	R4	N/A	R4	N/A	N/A
FHIR terminology module resources	+	+	+	N/A	+	-	+
FHIR terminology module operations	+	+	+	N/A	+	-	+
Internal data model	N/A	N/A	N/A	N/A	own data model	N/A	N/A
Multilingual terminology	-	+	-	-	+	-	-
Versioning	+	+	+	N/A	+	+	+
Web interface	+	+	+	+	-	+	+
Licence type	C	C	C	C	F	F	C

Legend: + means criteria is present in the Terminology Server; - means criteria is not present in the Terminology Server; N/A means no answer available for this criteria; C means commercial licence type; F means free licence type.

3.1.3 Key contributions

- **Framework for evaluation**: The paper provides a comprehensive framework for evaluating clinical terminology servers, which can be used by healthcare professionals, system developers, and researchers to make informed decisions (Table 5).
- **Comparison of existing servers**: It offers valuable insights into the strengths and weaknesses of existing terminology servers, highlighting areas for improvement.
- **Identification of gaps**: It highlights the limitations of current servers in supporting modern interoperability standards and multilingual content.
- **Proposal for a new solution**: It proposes the development of a new, enhanced terminology server (TermX) to address the identified gaps and meet the evolving needs of healthcare.
- **Promotion of open collaboration**: It emphasises the importance of open-source solutions and community-driven development to foster collaboration and transparency in clinical terminology management.

This paper is a contributes to the field of healthcare informatics, providing a clear roadmap for improving clinical terminology management and enhancing interoperability across healthcare systems.

3.2 Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia [III]

The study explores the transition of the ENHIS from a document-based approach using HL7 CDA to an event-based approach using HL7 FHIR.

The focus is on the migration process of the Infectious Disease Information System (NAKIS) in Estonia, particularly the socioeconomic status data of patients. The objective is to analyse and identify suitable FHIR resources and terminologies for representing socioeconomic status, examining existing FHIR profiles, openEHR clinical models, and terminologies such as SNOMED and LOINC about education, employment, and other socioeconomic information. A universal methodology for migrating HL7 CDA documents to FHIR resources is proposed, ensuring accurate and efficient data representation.

3.2.1 Problems addressed

The current document-based approach, relying on HL7 CDA in the ENHIS, faces limitations in timely information sharing and data interoperability. To address these challenges, there is a need to transition to an event-based approach using HL7 FHIR, which would enable more efficient and accurate data exchange, particularly for socioeconomic status, within the NAKIS. This transition involves breaking down complex data elements into simpler, more manageable components that can be updated frequently. To ensure accurate data representation and seamless interoperability, the data models and terminologies must be clearly documented and standardised.

3.2.2 Novelty

The study introduces a universal and reusable methodology for migrating HL7 CDA documents and their components to FHIR resources. It emphasises the importance of deconstructing complex data elements, using appropriate search keywords, and leveraging FHIR and openEHR standards, FHIR Implementation Guides, and standardised terminologies, such as SNOMED and LOINC. The research also highlights the benefits of using FHIR Observation resources (Figure 6) for dynamic and frequently changing data such as socioeconomic status.

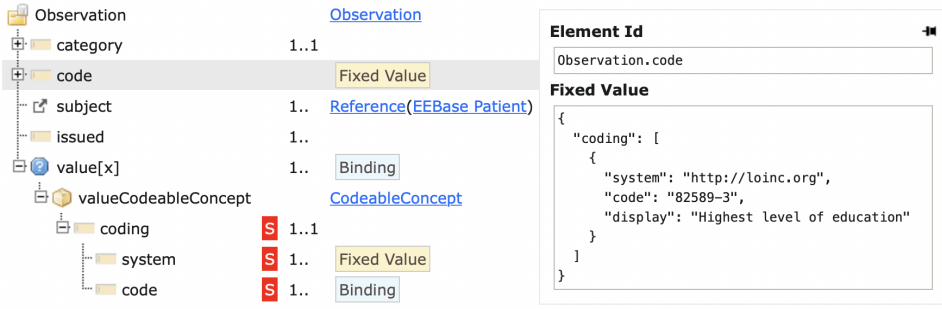

Figure 6: Component of socioeconomic status: Highest level of education

3.2.3 Key contributions

- **Framework for migration**: A detailed framework for migrating HL7 CDA documents to FHIR resources is provided, ensuring accurate data representation and interoperability, which can be applied to various healthcare information systems.
- **Socioeconomic status data**: By proposing the use of FHIR Observation resources for socioeconomic status, the representation and management of patient data are enhanced.
- **Interoperability and standardisation**: Leveraging existing FHIR profiles, openEHR clinical models, and terminologies ensures that the migrated data is interoperable and standardised, facilitating better data exchange and usage.
- **Practical implementation**: The paper demonstrates the practical application of the migration methodology in the NAKIS, validating its effectiveness.
- **Guidance for future implementations**: The methodology and practical steps outlined serve as a valuable guide for other healthcare systems undergoing similar transitions.

This study contributes to healthcare informatics by providing a clear and practical approach to transitioning from HL7 CDA to FHIR, with a focus on improving data representation and interoperability.

3.3 TermX: The semantic interoperability, knowledge management and sharing platform [I]

The study explores the process of development and implementation of TermX, an open-source platform designed to enhance semantic interoperability, knowledge management, and data sharing in healthcare systems. The TermX platform includes a terminology server, a Wiki, a model designer, a transformation editor, and tools for authoring and publishing.

TermX aims to improve access to terminology, simplify the design of data models, optimise data transformations across various models, and develop implementation guides through an intuitive web interface. The platform ensures open and standardised access to published terminology, data models, and schemas and follows the FHIR standard. TermX offers multilingual clinical terminology, resource descriptions, a visual editor for FHIR Mapping Language (FML) transformations, and a Wiki for knowledge management, making it a versatile tool for both technical IT professionals and non-technical healthcare staff. The platform's architecture is modular, allowing customisation and extension to meet specific needs.

3.3.1 Problems addressed

Achieving semantic interoperability in healthcare is challenging due to the complexity of managing and exchanging clinical data across diverse systems and standards. Existing tools often lack integration, multilingual support, and user-friendly interfaces and require extensive human expertise, making it difficult for healthcare professionals to develop and maintain terminologies, data models, and transformations. Implementing and maintaining FHIR standards is complex and resource-intensive. The main goal of this work is a comprehensive and clear description of the TermX software architecture, including the main components and their interactions, its design principles, and its practical implementation, which will provide a thorough understanding of the software, making it valuable for those interested in the technical aspects of the software.

3.3.2 Novelty

TermX introduces several innovative features and approaches to address these challenges:

- **Comprehensive platform**: Unlike many existing solutions that focus on specific tasks, TermX offers a comprehensive suite of tools for terminology management, data modelling, transformation, and publishing, all within a single platform.
- **Visual FML editor**: The visual editor for FML transformations is a unique feature that simplifies the creation and management of data transformations, making it accessible to business analysts without needing deep technical expertise and reducing the complexity of the FML language.
- **Enhanced data transformation**: It supports the transformation of any data format to any other data format, with a focus on FHIR transformations, ensuring flexibility and adaptability.
- **Knowledge management**: It includes a Wiki for collaborative knowledge management, facilitating the creation and sharing of well-formatted content, a thesaurus, and implementation guides.

- **Advanced publishing capabilities**: It enables the bidirectional synchronisation of resources with external storage, the generation of implementation guides, and static sites, enhancing information sharing.
- **Integration with external services**: TermX supports integration with various external services, including GitHub for synchronising terminology resources, third-party FHIR-compatible terminology servers, and the IHTSDO Snowstorm server for SNOMED management.
- **Modular architecture**: It features a modular architecture that allows customisation and extension, enabling users to tailor the platform to their specific needs.
- **Customisability and extensibility**: The platform is highly customisable and can be extended to meet specific customer needs. Modules can be replaced or enhanced with customised implementations, and new features can be added through plugins or modules.
- **Open-source and no vendor lock-in**: TermX is available under the MIT licence, ensuring no vendor lock-in and promoting widespread adoption and collaboration.

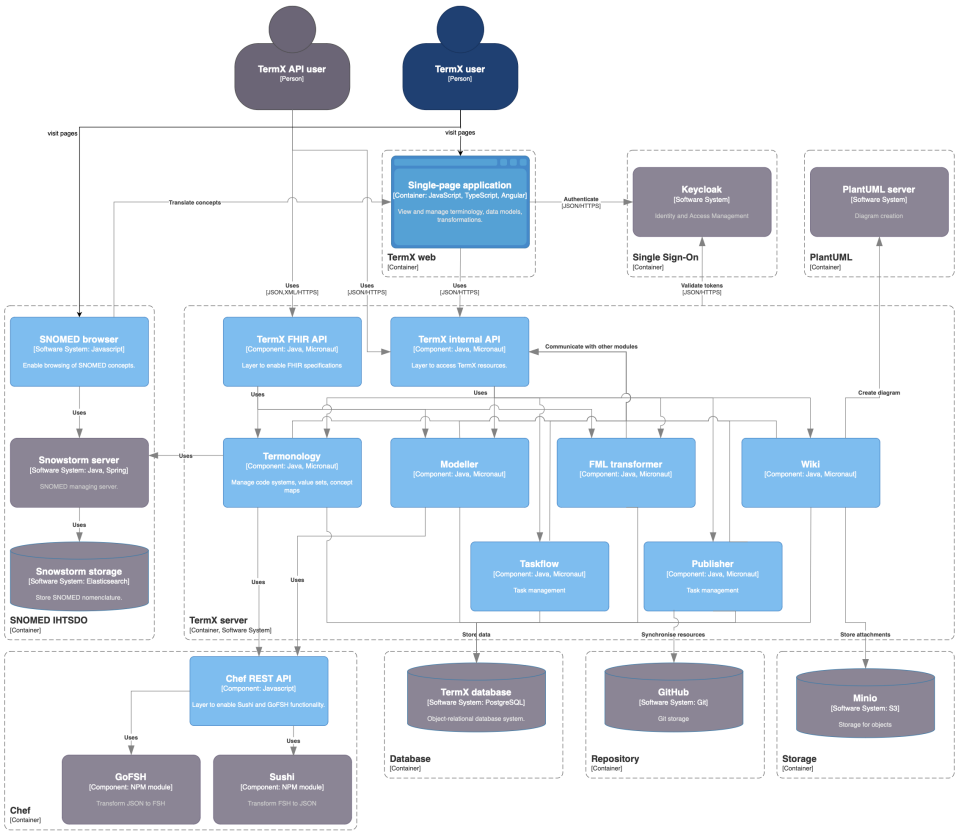

Figure 7: TermX component diagram

3.3.3 Key contributions

- **Platform for interoperability**: TermX offers a comprehensive platform for achieving interoperability in healthcare information systems.
- **Functional overview**: The practical applications and benefits of TermX are highlighted through detailed descriptions of functional modules and illustrative examples.
- **Architecture overview**: Understanding of the TermX operating principles will be improved by demonstrating software architecture, including internally developed micro-services, and external components, such as PostgreSQL, Minio, GitHub, Keycloak, PlantUML, Snowstorm server, and others, will improve understanding of the TermX operating principles (Figure 7).
- **Impact**: TermX reduces the time required to learn the basics of terminology, modelling, and transformations, optimises human resources, and simplifies the adaptation of standards.

This study contributes to advancing healthcare interoperability by providing a comprehensive, customisable, and open-source platform that enhances semantic interoperability, simplifies data model design and transformations, supports multilingual clinical terminology, and integrates various external services, thereby improving data management practices and fostering collaboration across healthcare systems.

3.4 Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components [II]

The article introduces a tool and techniques for achieving health data semantic interoperability by transforming HL7 CDA to FHIR using reusable visual components. It aims to simplify complex health data transformations, making them accessible to domain experts with minimal technical skills.

The study presents a tool and techniques developed using the design science methodology to facilitate the interoperability of EHRs by enabling the seamless unification of various health data formats in real time. The tool simplifies complex health data transformations, allowing domain experts to specify and validate intricate data transformation rules and maps. This approach addresses the ongoing transition of the ENHIS from HL7 CDA to FHIR, but it is general enough to be used for other data transformation needs, including the EHDS ecosystem.

3.4.1 Problem addressed

There is a need for semantic interoperability of health data in various formats. The ENHIS is transitioning from CDA format to FHIR. To mitigate the risks associated with data migration, the system must operate with legacy CDA data while storing new data in FHIR format, necessitating on-the-fly semantic interoperability between both formats. Additionally, there is the broader problem of transforming EHR data from one format to another in a semantically interoperable manner.

3.4.2 Novelty

The study introduces a novel approach using reusable visual transformation components to create and validate transformation rules and maps (Figure 8). This method allows domain experts with minimal technical skills to specify and validate data transformation

rules, making the process more efficient and accessible. The integration of these components into a unified tool simplifies the transformation of complex health data formats, enhancing clarity and promoting reuse.

Figure 8: Example of the reusable transformation component for CDA to FHIR

3.4.3 Contributions

- **Development of a tool for semantic interoperability**: A tool that facilitates the transformation of health data from CDA to FHIR using reusable visual components was created.
- **Simplification of data transformation**: Techniques that allow domain experts to specify and validate data transformation rules and maps with minimal technical knowledge were implemented.
- **Reusable transformation components**: Reusable visual transformation components that enhance the efficiency and reliability of data transformations were developed.
- **Validation and collaboration**: The tool and techniques were validated in collaboration with domain experts from various countries, demonstrating their usability and effectiveness.
- **Support for federated interoperability**: The tool supports federated semantic interoperability, making it applicable to various health data transformation needs, including the EHDS ecosystem.

This paper contributes to the advancement of health data semantic interoperability by providing a user-friendly tool that enhances productivity and communication across healthcare systems.

3.5 Modelling a patient identifier system in the Estonian National Health Information System [V]

The paper presents the development and assessment of a coding system for patient identifiers within the ENHIS. It aims to enhance the precision of health records and provide a robust method for identifying patients from diverse backgrounds.

The study investigates the design of a patient identifier system in the ENHIS during its transition from HL7 V3 to HL7 FHIR communication protocols. This transition involved evolving from an object identifier (OID) system to a uniform resource locator (URL) system. The research introduces an IdentifierDomain coding system tailored to patient identification, which is user-friendly, semantically clear, backward compatible with the OID system, and expandable and aligns with FHIR standards. The proposed system was tested using examples from the Estonian Patient Register and validated by key stakeholders.

3.5.1 Problem addressed

Accurate patient identification is crucial in healthcare to prevent fatal errors caused by mistaken identity. While identifying citizens is generally well-regulated, identifying foreign, unknown, or anonymous patients is more challenging and often lacks sufficient regulation. The ENHIS system faced issues with varied practices among healthcare institutions for identifying foreign patients and the technical limitation of using only one patient ID per document.

3.5.2 Novelty

The study introduces a novel URL-based identifier domain that replaces the less user-friendly OID system. This new system is designed to be human-readable, flexible, and capable of supporting multiple identifiers for patients, including residents, foreigners, and unknown patients. The methodology allows for the automated issuing of identifier systems and supports the seamless integration of new identifier types and countries (Figure 9).

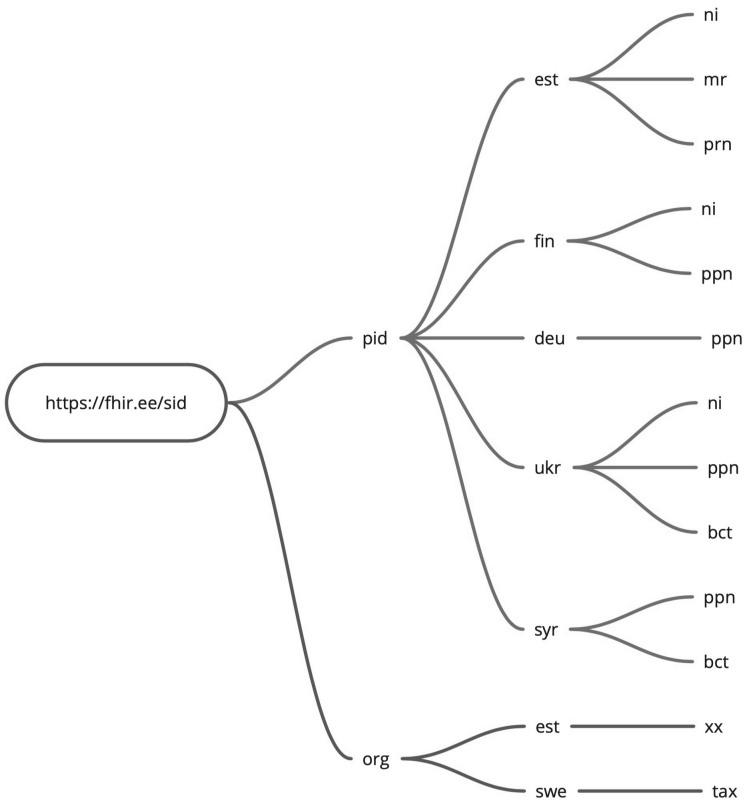

Figure 9: Visualisation of patient identifier domain types and their subcategories

3.5.3 Key contributions

- **Development of a URL-based identifier domain**: A human-readable, flexible, and expandable code system with a TermX terminology server was created.
- **Enhancement of patient identification**: The accuracy and reliability of patient identification in the ENHIS were improved by supporting multiple identifiers and linking patient records.
- **Support for foreign and unknown patients**: A robust method for registering and identifying foreign and unknown patients was developed, addressing a significant gap in the existing system.
- **Validation and implementation**: The proposed system was validated with key stakeholders and successfully integrated into the FHIR-based Estonian Master Patient Index (MPI).
- **International applicability**: A methodology that can be generalised for use in other national and regional health information systems was provided, promoting global interoperability.

This paper contributes to the advancement of patient identification systems by providing a robust solution that enhances data quality and interoperability in healthcare.

4 TermX

TermX is an open-source knowledge management and sharing platform, including a terminology server, a Wiki, a model designer, a transformation editor, and tools for authoring and publishing [1].

4.1 Software functionality

4.1.1 Terminology server

The terminology server is based on CTS2 [59], a standard for accessing and managing terminologies. TermX empowers users to create, store, update, and query terminologies from diverse clinical and administrative domains and multiple human languages. TermX supports the FHIR terminology module, including code systems (Figure 10), code system supplements, value sets, concept maps, and naming systems. TermX also supports the versioning, validation, and authoring of terminologies. The TermX SNOMED module enables the management and translation of SNOMED concepts.

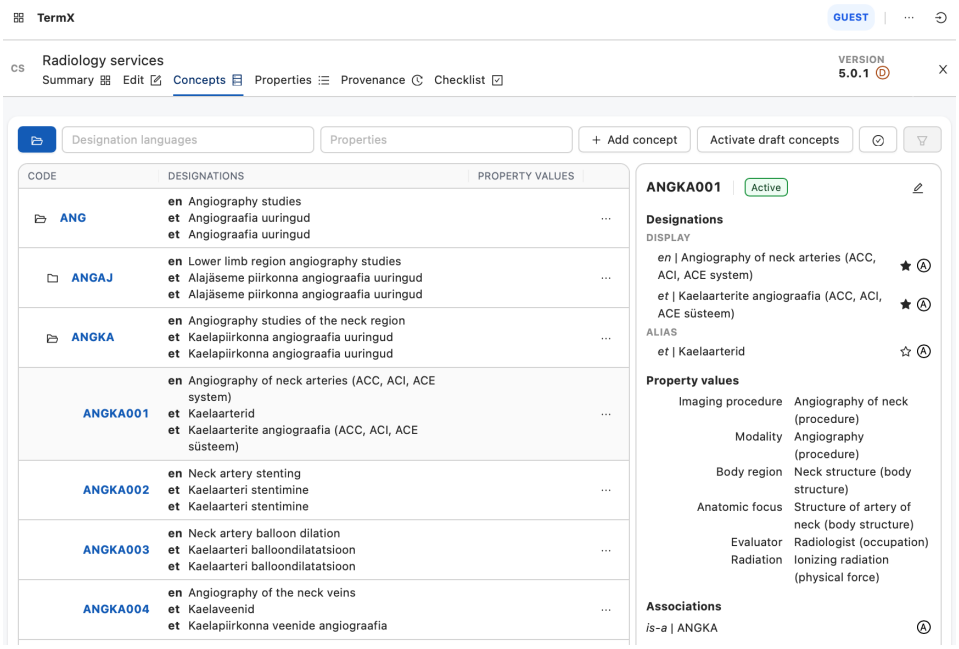

Figure 10: Snapshot of the TermX user interface with the hierarchy of radiology services and details of the concept ANGKA001

4.1.2 Model Designer

The Model Designer implements the FHIR StructureDefinition specification [62], which prescribes rules for logical data models, FHIR resources, and profiles as well as their elements and data types. With Model Designer, creating and managing data models through a user-friendly interface is possible. For FHIR StructureDefinition, both JSON and FSH notations are supported. The model designer capabilities allow the import/export of models and define attributes, data types, cardinality, and terminology binding, among other features.

4.1.3 Visual editor for FML data transformations

The FHIR Mapping Language (FML) is a transformation language specifically designed to transform HL7 FHIR resources to/from alternative representations [63], including different logical data models [64], FHIR resources, CDA documents [65], etc. TermX provides a unique visual FML editor as a designer of explicitly designed FML transformations for business analysts (Figure 11). The objective of the FML editor is to visually represent transformations, hide the complexity of the FML language, and facilitate rapid adaptation to it.

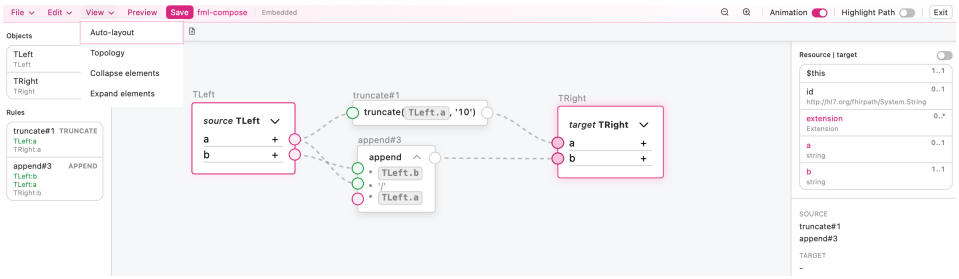

Figure 11: FML transformation of TLeft model to TRight with text functions

4.1.4 Wiki

A Wiki is a web-based collaborative platform that enables well-formatted content creation and editing in an organised manner. It serves as a foundation for knowledge management, including a thesaurus, tutorials, and custom pages for static sites and FHIR implementation guides. The thesaurus plays a crucial role in defining terms for effective communication. Additionally, Wiki plugins allow the easy referencing of terminology, models, and transformations and their inclusion on pages. Markdown syntax is commonly used to edit Wiki pages. Wiki implementation is based on Markdown-it processors [66].

4.1.5 Publisher

A publisher facilitates the organisation of terminology, models, and Wiki pages into logical spaces. It enables the bidirectional synchronisation of resources with external storage. Specifically, the publisher supports synchronisation with GitHub repositories and syndication with terminology servers that adhere to the FHIR Terminology API. Additionally, it provides the capability to generate implementation guides and static sites.

4.2 Software architecture

TermX consists of the micro-services required to run TermX. The TermX micro-services include (Figure 7): 1) **TermX server**, the main back-end application, which was written in Java and is responsible for orchestrating TermX modules; 2) **TermX web**, an Angular-based front-end application that communicates with the server via a RESTful API and offers a user-friendly interface for effortless task execution; 3) **Chef**, the wrapper for Mitre Sushi and GoFSH projects [67], which provides the ability to transform the FHIR JSON notation of the resources into Shorthand FSH notation[45] and vice versa. TermX web includes **FML editor** and FHIR **StructureDefinition viewer**, which are Javascript- and Angular-based libraries for data transformation management. These libraries interact with StructureMap [68] resources while implementing the FHIR FML [69, 63] specification.

The TermX server is composed of independent modules. Communication between these modules is organised through the 'termx-api' module, which provides communication interfaces. Back-end services are available through the internal OpenAPI REST API and the FHIR API. The implementation of the modules varies based on the tasks they perform, and they communicate with external micro-services as needed. The persistence layer utilises the relational and non-SQL features of the PostgreSQL database server. Database has an isolated schema for every module. The terminology module, which employs the CTS2 specification, describes a variety of relatively fine-grained resources and associations. These are implemented as an Entity-Relationship model in the database (Figure 12) [60].

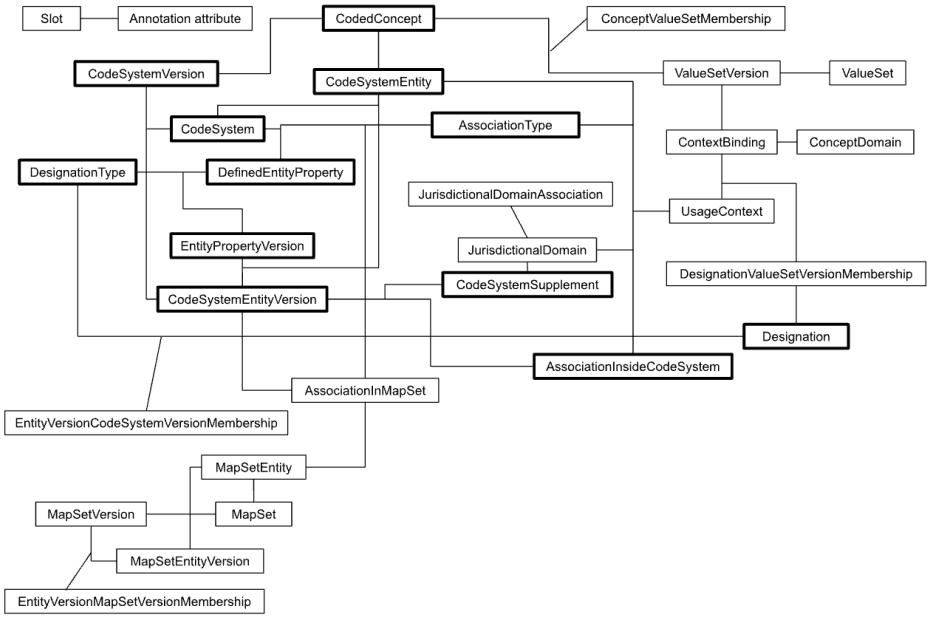

Figure 12: The entities of the CTS2 conceptual model

The external components include (Figure 7): 1) the *PostgreSQL database*, which is used as internal data storage in TermX; 2) a Minio-based [70] *binary object repository* for storing attachments and large files; 3) *GitHub* for synchronising terminology resources and Wiki pages and automating processes related to the FHIR implementation guide and static site generation; 4) third-party *FHIR-compatible terminology servers* that may be used as sources or destinations for the synchronisation of terminology resources; 5) the IHTSDO *Snowstorm* server and SCT browser [71] used for the management of SNOMED terminology; 6) the Keycloak *OpenID Connect server* used for user and role management [72]; and 7) the *PlantUML server* used to generate images of UML diagrams [73]. The adaptable and flexible TermX architecture allows integration into existing customer ecosystems. TermX is an open-source platform available under an MIT licence [74]. TermX components are available as source code, libraries, and Docker images [75]. The deployment configurations for Docker Compose and Helm simplify the installation process [76].

4.3 Actual use

The terminology server is the most popular component of TermX. The TermX terminology module has been adopted by the national standardisation agencies of Estonia (TEHIK) and Lithuania (LMB) for terminology development, management and publishing. Conversely, in academic projects and research involving TalTech, Model Designer, transformation tools, and Wiki for documentation are more prevalent.

Users tend to prefer TermX due to the relative simplicity of its web interface, the ability to customise this interface, and its integration with their internal software products. Other appealing features include enterprise single sign-on (SSO), the option of on-premise installation, and the absence of vendor lock-in, thanks to the open MIT licence.

5 Discussion

5.1 Outline of research findings and evaluation

In the healthcare sector, significant attention is devoted to interoperability issues. Over time, various standards have been established for data models (HL7 RIM, CTS2, OpenEHR Reference Model), terminology (SNOMED, LOINC), document representation (CDA, FHIR Composition), data exchange (FHIR), and data transformation (QVT, FML) (see tables 1 and 4).

However, the interoperability problem remains unresolved as a whole [VI].

We designed a framework (Section 2.2) to focus our research on the most popular modelling, terminology, communication, data representation, and encoding standards. We specified the target group, which includes business and information analysts and architects responsible for process coordination, defining and coding data models and terminology, and describing the integration rules for applications into healthcare systems.

There is a lack of high-quality open-source tools that offer proper user interfaces and multilingual support [I], [IV]. Additionally, there is a lack of modelling tools [VI] and transformation editors [II]. Existing terminology servers exhibit several deficiencies [IV], and the development of implementation guides is notably complex [III]. Furthermore, many tools necessitate highly advanced programming skills [VIII].

To overcome this, we propose an interoperability platform, TermX, which improves the adoption of the standards described in the framework (Section 2.2), such as FHIR and CTS2. The development was initiated with the prototype of the terminology server [60]. After the successful implementation of the prototype, the research grant was requested and received. Expert knowledge, gathered through interviews with a nine-member focus group of international eHealth experts (primarily HL7 experts), TalTech researchers, and the practical experience of the author from designing national and hospital information systems, helped clarify the requirements for the interoperability platform. The TermX platform was created [77, 51], documented [78], and published as open-source software [74] (publications [I] and [IX]). TermX incorporates a terminology server, a Wiki, a model designer, a transformation editor, and tools for authoring and publishing.

The developed modules of the TermX platform underwent evaluation by a group of experts to assess their practical usefulness and compliance with established standards. Subsequently, evaluation phases were conducted for the terminology, modelling, and transformation modules (publications [II], [V], and [X]).

At the same time, the terminology server was thoroughly verified and deployed for production use at the LMB. Later, it was adopted by TEHIK for similar purposes. The terminology server is planned to be used in Uzbekistan.

5.2 Summary of related work

Interoperability frameworks provide a structure along with methods, requirements, and qualification criteria but typically do not prescribe specific standards and techniques for achieving results [32].

Based on the descriptions of the informational and applicational layers in the ReEIF, semantic interoperability can be ensured through a shared data model and terminology management as well as the adaptation and implementation of terminology and data exchange services [16].

5.2.1 Terminology

The basis of modern terminology theory emerged in 1930 with the work of Eugen Würster [79]. Wüster developed a theory of terms and concepts that later became entrenched as the terminology standard promulgated by ISO [80, 81]. In 1998, James Cimino introduced a set of desiderata that must be satisfied by medical terminologies for them to support modern computer applications [82, 83]. The principles stated in Cimino's desiderata are still the main criteria used to assess the consistency of clinical terminology today and incorporated in the development of terminologies such as SNOMED CT [84], UMLS [85], NHS Clinical Terms [21], and terminology Checklist [61].

In the field of terminology servers, the GALEN terminology server can be singled out [86]; it was well described back in the 1990s. By now, his ideology and solution are hopelessly outdated. We can note the theoretical works [87, 88], but not the implementations. Alternatively, there are commercial servers, such as Ontoserver [43], but no access to the code and a limited amount of research work.

In recent years, the specifications of the HL7 FHIR terminology module and HL7 Common Terminology Services (CTS2) standard were released. Despite a robust theoretical foundation in the field of terminology, prior to the advent of TermX, there was no open-source server specifically designed and developed to support both standards [60].

5.2.2 Data modelling

The foundational theories of data modelling are attributed to Edgar Codd, who developed the theoretical groundwork for a relational model for database management systems in 1969 [89]. Today, the Object Management Group (OMG) [90] plays a crucial role by developing and maintaining data modelling standards such as the Unified Modeling Language (UML) and the Meta-Object Facility (MOF). Model-driven architecture (MDA) is an approach to software design and development that utilises MOF, UML, and platform-independent models (PIMs) to drive the development process and ensure that models accurately represent data and business logic, thereby enhancing the interoperability, reusability, and maintainability of software systems [91, 92].

The FHIR StructureDefinition resource was developed by HL7 to describe the structure of other FHIR resources, data types, extensions, and logical models [62, 93]. Essentially, it lays out a set of data element definitions and their associated rules of usage. TermX provides a user-friendly interface for the creation and management of such data models [I].

5.2.3 Transformations

The concept of 'mapping language' (or data transformation language) lies in establishing a platform-independent specification that can be implemented across various programming languages [94]. Model-to-model transformations are typically articulated in specialised domain-specific languages, often known as model transformation languages (MTLs) [95]. 'Query/View/Transformation' (QVT) is a specification developed by the OMG to describe transformation rules between different data models in the MDA domain [96]. FML [63] is a relatively new, bespoke transformation language specifically designed to transform HL7 FHIR resources into alternative representations [64]. The FHIR Management Group created the mapping language as a specification of the QVT framework[97] for model-transformation languages.

Our research has identified that mapping languages are essential for defining mappings and transformations between various data structures. The QVT framework is widely used in model-driven development, particularly for managing model transforma-

tion tasks. FML is the sole mapping language that fully implements the QVT specification and is endorsed by HL7 within the healthcare sector. Although the functionality of FML is somewhat limited [51], it is sufficiently robust to handle complex transformations, such as those from CDA to FHIR [II].

5.2.4 Knowledge sharing and publication

The implementation guide (IG) represents a relatively new domain for disseminating the outcomes of standardisation efforts, which remains largely unexplored in academic literature. FHIR implementation guidelines (IG) were popularised in 2019. The FHIR community has developed numerous guidelines and profiles [98, 99], yet there are only a few scientific papers [93] that primarily focus on the results rather than the underlying processes and methodologies. Nevertheless, IGs are critical, as they constitute the final specifications for implementation and are utilised by information analysts, developers, and healthcare professionals. Future research into IGs could address the existing gap in this domain.

5.2.5 Problem relevance

The body of literature on FHIR is extensive, encompassing both the standard itself [21, 40] and various FHIR-related projects [69, 22]. The FHIR standard, along with its associated tools, is in a state of continuous evolution. A plethora of open-source libraries is available for developers [100]. Initially, most FHIR servers were open source; however, many have since been commercialised (e.g. Spark, Aidbox) or discontinued (e.g. IBM LinuxForHealth FHIR server).

In the realm of terminology servers, open-source solutions are conspicuously scarce. Snowstorm stands as a notable exception, though it is tailored specifically for the management of the SNOMED nomenclature. While tools for data modelling and transformations are available in other domains or for different healthcare standards, such tools are notably absent for FHIR.

The tooling supporting the HL7 FHIR standard remains largely unexplored in academic literature. Several factors contribute to this gap. First, commercial entities are reluctant to disclose proprietary information that provides them with a competitive edge. Second, academic institutions are not typically involved in the development of these tools, resulting in the architecture and algorithms of such tools remaining obscure to the academic community.

Therefore, it is imperative to develop comprehensive tools that can bridge these gaps, fostering greater transparency and collaboration between commercial and academic sectors. This will not only enhance the utility of FHIR but also ensure its robust implementation across diverse healthcare settings.

TermX, developed as part of this doctoral research, aims to not only address these identified gaps but also elevate interoperability to a new level. It is designed to be an accessible tool for non-technical personnel, thereby broadening its usability and impact.

5.3 Summary of contribution

The aim of this thesis (RQ) was to develop a solution accessible to non-technical personnel, specifically information and business analysts, to achieve semantic compatibility in electronic health records in alignment with international healthcare standards. This enquiry was divided into two primary subtopics: (sub-RQ1) terminology management; and (sub-RQ2) data model management and transformations between models. In the third (sub-RQ3), we aimed to assess the feasibility and effectiveness of our proposed solution.

The **main contribution** of our research is the creation of TermX, an open-source plat-

form designed to improve interoperability among healthcare institutions and systems [77]. TermX enhances the interoperability of both the primary and secondary use of health data by facilitating terminology unification, streamlining data model design, and supporting data transformations between models and communication protocols.

The essential components of TermX include the **Terminology Server**, the **Model Designer**, the visual **FML Editor**, and a **Publisher** that facilitates structuring terminology, models, and Wiki pages into logical spaces (see Section 4.1).

Informational analysts, as domain experts, possess extensive knowledge of the ontology, models, and terminology pertinent to their field. They are tasked with planning and ensuring the accuracy of information design. TermX is a robust tool specifically designed to support analysts in these tasks. Its visual interface enhances clarity, facilitates the reuse of terminology, models, and transformation components, and abstracts the complexities of the FHIR framework, FML mapping language, and interoperability standards, thereby enabling analysts to quickly adapt to its use.

Two design-science cycles were employed to test the primary artefact of this thesis, TermX. The first cycle validated the terminology server by developing a new radiology classifier. The second cycle validated both the modeller and FML editor by designing transformations from CDA to FHIR. Finally, TermX was validated in production environments in Lithuania and Estonia.

TermX has changed its users' daily practices in several ways. It enables the precise definition and management of medical terms, ensuring the quality, consistency, and interoperability of terminologies and data across healthcare systems. The platform's intuitive web user interface and data exchange features with external services have improved the efficiency of terminology and data model definition and management. Furthermore, the visual editor of data transformation hides the complexity of the FML language, allowing for wider adoption. In addition, TermX reduces the time required to learn the basics of terminology, optimises human resources, and simplifies the adaptation of the standards.

The implementation of TermX can enhance both the primary and secondary use of health data. For primary use, which involves direct patient care and data entry into information systems, TermX can provide the necessary terminology and data models to be used during data input and validation, ensuring the entry of high-quality data. Even if the TermX data model is not utilised during primary data collection, such models are typically employed for healthcare data exchange within data spaces. For instance, the ENHIS is an example of a data space that connects all the healthcare institutions in Estonia. Every data space should specify the supported data models and terminology. Improving primary data usage through well-documented models and terminology leads to enhanced secondary data usage. When the same models and terminology are used at both primary and secondary levels, data transformation is unnecessary, and data quality is ensured from the point of collection. In cases where there are differences in models, terminology, or their versions, transformations can be used to convert the data accordingly.

For secondary use, which includes research, public health, and policymaking, standardised healthcare models and terminology provided by TermX, along with data collected according to these standards, can enable researchers to conduct more robust studies, identify trends, and generate insights that can inform public health strategies and healthcare policies.

In summary, TermX has the potential to transform the landscape of health data utilisation, making it more efficient, accessible, and impactful across both primary and secondary data use.

An *additional contribution* of this thesis lies in the methods and outcomes presented

in the associated papers. The publication [III] served as a foundational reference for analysing the ENHIS data composition for migration to FHIR [53]. The developed Patient Identifier Domain [V] is currently utilised in the production environment of the ENHIS Patient Registry. The Checklist [61] has contributed to enhancing the quality of ENHIS terminology. Additionally, the CDA to FHIR prototype (publication [II]) has been adopted as the basis for the ENHIS solution.

A *side contribution* of this thesis is its facilitation of academic research, with TermX serving as a catalyst for new topics. Several academic studies involving TalTech MSc students were initiated to evaluate sub-RQ1 and sub-RQ2. Ongoing research projects with MSc and BSc students include FHIR to OMOP transformation, visualisation of FHIR StructureDefinitions, and the design of UML notation for FHIR StructureDefinitions. Several doctoral studies using TermX are also planned. Additionally, TermX has been incorporated into the TalTech healthcare interoperability course.

The results presented in this section demonstrate full coverage of the research questions.

5.4 Limitations and implications for further research

There are numerous interoperability standards, and it is not feasible to address all of them within the scope of this thesis. This thesis focuses on the latest modelling and terminology standards and frameworks from the HL7 family, including HL7 FHIR, HL7 CTS2, HL7 CDA, the FML language, as well as terminology services and implementation guide technologies. This thesis does not cover other important topics, such as communication, security, privacy, and safety. As mentioned in the previous section, several research studies that aim to expand the range of included standards and address additional challenges are currently underway and planned.

Healthcare interoperability standards, frameworks and libraries are under continuous development. With the development of a software product comes the responsibility for its further development and compliance with the latest developments. TermX users are interested in the new features and improved usability of the web interface.

The current solution is optimally designed for a standalone setup with a single master organisation. One of the forthcoming tasks is to develop a portal that supports a multi-tenant environment with multiple master organisations. Another direction involves leveraging the support of TalTech students and researchers to develop a well-known library of applications and libraries, ensuring tight integration with the HL7 community and promoting TalTech's research and solutions.

6 Conclusion

The journey towards achieving semantic interoperability in primary and secondary health data use is complex and has a significant impact on the healthcare domain, given the vast array of data standards, evolving regulatory requirements, and the diversity of healthcare systems worldwide. Despite extensive research and the development of complex standards, their implementation is often hindered by their complexity and the steep learning curve they present. Traditional interoperability solutions have typically required extensive technical knowledge, limiting their adoption to specialised users and creating bottlenecks in healthcare data sharing. To address these challenges, it is essential to simplify the adoption of these standards and make them more accessible to business professionals. Tools such as TermX serve as vital bridges, facilitating the understanding and application of modern complex standards through visualisation and user-friendly interfaces.

In this thesis, we address gaps by developing TermX, a comprehensive interoperability platform designed to simplify healthcare data sharing across systems and national borders, making it accessible to non-technical users. TermX offers a suite of modular tools — including a terminology server, model designer, and data transformation editor — which empower analysts to manage and adopt complex healthcare standards like FHIR without extensive programming expertise. The research leveraged the design science (DS) methodology to iteratively design, create, test, and refine artefacts, ensuring it met both theoretical and practical objectives in healthcare informatics.

The real-world deployment of TermX in healthcare institutions in Estonia and Lithuania illustrates its efficacy. These implementations demonstrated TermX's capacity to streamline data model development and terminology management, reduce deployment costs, and simplify adherence to interoperability standards. Through its support for FHIR API and multilingual capabilities, TermX proved itself a viable solution for healthcare systems aiming to enhance both primary and secondary data uses in electronic health records (EHRs). This adaptability enables healthcare providers and policymakers to foster more cohesive data-sharing environments and build infrastructures that are future-ready.

The key contributions of TermX include a substantial reduction in the learning curve for interoperability standards and the technical and financial barriers associated with training and onboarding personnel. The platform's transformation editor, capable of executing intricate data translations (e.g. from HL7 CDA to FHIR), is especially beneficial in environments where data continuity and standardisation are essential for patient care and regulatory compliance. Furthermore, the platform's terminology management tools align with international frameworks, enabling the consistent application and evaluation of medical terminologies.

Looking forward, future directions for TermX include expanding its functionality to support multi-tenant environments, thereby enabling broader use across healthcare organisations. Further enhancements will explore integration with emerging interoperability standards and technologies, aligning TermX with evolving global interoperability initiatives such as the European Health Data Space (EHDS) and WHO SMART guidelines. Integrating TermX into a broader ecosystem of interoperable tools can pave the way for a more unified global approach to healthcare data exchange.

In summary, this thesis has presented TermX as an effective, scalable solution to address the challenges in healthcare interoperability. This tool addresses critical business and social needs by enabling domain experts to develop, manage, and publish terminology, data models, and transformations. By bridging the gap between complex technical standards and user accessibility, TermX enhances the usability and adaptability of electronic health data, empowering non-technical users to actively participate in interoper-

ability initiatives. It supports the efficient and accurate management of health data and the coordination of healthcare processes. This platform not only reduces the barriers to interoperability but also sets the foundation for more efficient, standardised healthcare systems that can ultimately improve patient care and data accessibility.

List of Figures

List of Tables

References

[1] Bossenko, Igor and Piho, Gunnar and Ivanova, Marina and Ross, Peeter, "TermX: The semantic interoperability, knowledge management and sharing platform," *SoftwareX*, vol. 27, 2024.

[2] I. Bossenko, R. Randmaa, G. Piho, and P. Ross, "Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components," *Frontiers in Digital Health*, p. 30, 2024.

[3] I. Bossenko, K. Linna, G. Piho, and P. Ross, "Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pp. 882–885, 2023.

[4] M. Ivanova, I. Bossenko, and G. Piho, "Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution," *Lecture Notes in Business Information Processing*, vol. 531, p. 12, 2024.

[5] I. Bossenko, G. Piho, and P. Ross, "Modelling a patient identifier system in the Estonian National Health Information System," *Lecture Notes in Business Information Processing*, vol. 531, p. 14, 2024.

[6] P. Ross, J. Metsallik, K. J. I. Kankainen, I. Bossenko, C. Mäe, and M. Maasik, "Health Sense: development of a universal data model and a standard for continuity of treatment paths based on international standards of new generation health information systems," *TalTech Digikogu*, 2023. English translation: https://zenodo.org/records/14599236.

[7] I. Bossenko, G. Piho, and P. Ross, "Forward and Backward Compatibility Design Techniques Applying the HL7 FHIR Standard," in *HEDA@ Petri Nets*, 2022.

[8] R. Randmaa, I. Bossenko, T. Klementi, G. Piho, and P. Ross, "Evaluating business meta-models for semantic interoperability with FHIR resources," in *HEDA@ Petri Nets*, 2022.

[9] Bossenko, Igor and Piho, Gunnar and Ross, Peeter, "TermX: A game changer in the healthcare interoperability," in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, vol. 316 of *Studies in Health Technology and Informatics*, pp. 88–89, IOS Press, 2024.

[10] M. Marquis, I. Bossenko, and P. Ross, "RadLex and SNOMED CT Integration: A Pilot Study for Standardising Radiology Classification," *Insights into Imaging*, p. 12, 2025.

[11] D. G. Katehakis and A. Kouroubali, "A framework for ehealth interoperability management," *Journal of Strategic Innovation and Sustainability*, vol. 14, no. 5, 2019.

[12] K. M. McDonald, V. Sundaram, D. M. Bravata, R. Lewis, N. Lin, S. A. Kraft, M. McKinnon, H. Paguntalan, and D. K. Owens, "Definitions of care coordination and related terms," in *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 7: Care Coordination)*, Agency for Healthcare Research and Quality (US), 2007.

[13] J. Weber-Jahnke, L. Peyton, and T. Topaloglou, "ehealth system interoperability," *Information Systems Frontiers*, vol. 14, pp. 1–3, 2012.

[14] M. Sprenger, V. van Pelt, and M. Seven, "Refinement of the eHealth European Interoperability Framework (ReEIF)." https://health.ec.europa.eu/document/download/84839361-c768-4ad5-b7ca-e46b2123fef2_en, 2015.

[15] R. H. Dolin and L. Alschuler, "Approaching semantic interoperability in health level seven," *Journal of the American Medical Informatics Association*, vol. 18, no. 1, pp. 99–103, 2011.

[16] B. H. de Mello, S. J. Rigo, C. A. da Costa, R. da Rosa Righi, B. Donida, M. R. Bez, and L. C. Schunke, "Semantic interoperability in health records standards: a systematic literature review," *Health and technology*, vol. 12, no. 2, pp. 255–272, 2022.

[17] T. Scheplitz, "Ensuring socio-technical interoperability in digital health innovation processes: An evaluation approach.," in *HEALTHINF*, pp. 264–275, 2022.

[18] H. Pirnejad, R. Bal, A. Stoop, and M. Berg, "Inter-organisational communication networks in healthcare: centralised versus decentralised approaches," *International Journal of Integrated Care*, 2007.

[19] T. Schliemann, C. Danielsen, T. Virtanen, R. Vuokko, G. A. Hardardottir, M. A. Alsaker, B. Aksnes, N. Eklöf, and E. Ericsson, *eHealth standardisation in the Nordic countries:: Technical and partially semantics standardisation as a strategic means for realising national policies in eHealth*. Nordic Council of Ministers, 2019.

[20] R. Egelkraut, B. Traxler, E. Helm, O. Krauss, and A. Schuler, "Open Infrastructure for Standardization of HL7® FHIR® Implementation Guides in Austria.," in *dHealth*, pp. 221–223, 2022.

[21] T. Benson and G. Grieve, *Principles of Health Interoperability*. Springer, 2021.

[22] M. Mercorella, M. Ciampi, M. Esposito, A. Esposito, and G. De Pietro, "An architectural model for extracting FHIR resources from CDA documents," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 597–603, IEEE, 2016.

[23] T. J. Eggebraaten, J. W. Tenner, and J. C. Dubbels, "A health-care data model based on the HL7 Reference Information Model," *IBM Systems Journal*, vol. 46, no. 1, pp. 5–18, 2007.

[24] HL7, "FHIR overview - Architects.." https://www.hl7.org/fhir/overview-arch.html#principles, 2024. Accessed: 2024-12-22.

[25] M. A. Kramer, "Reducing FHIR "Profiliferation": A Data-Driven Approach," in *AMIA Annual Symposium Proceedings*, vol. 2022, p. 634, American Medical Informatics Association, 2022.

[26] F. Bathelt, "The usage of ohdsi omop–a scoping review," *Proceedings of the German Medical Data Sciences (GMDS)*, pp. 95–95, 2021.

[27] S. Turbow, J. R. Hollberg, and M. K. Ali, "Electronic health record interoperability: how did we get here and how do we move forward?," in *JAMA Health Forum*, vol. 2, pp. e210253–e210253, American Medical Association, 2021.

[28] HIMSS, "Interoperability in Healthcare." Accessed: 2024-12-22.

[29] IHE, "About IHE." https://www.ihe.net/about_ihe/faq, 2024. Accessed: 2024-12-22.

[30] K. Baumann, O. Corcho, M. T. Horsch, T. Jouneau, M. Molinaro, S. Peorni, A. Scharnhorst, S. Vancauwenbergh, and L. Vogt, *Task Force Charter: Semantic Interoperability: Charter of the TF Semantic Interoperability*. EOSC, 2021.

[31] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[32] M. Sadeghi, A. Carenini, O. Corcho, M. Rossi, R. Santoro, and A. Vogelsang, "Interoperability of heterogeneous systems of systems: from requirements to a reference architecture," *The Journal of Supercomputing*, vol. 80, no. 7, pp. 8954–8987, 2024.

[33] WHO, "SMART Guidelines." https://www.who.int/teams/digital-health-and-innovation/smart-guidelines, 2024. Accessed: 2024-12-22.

[34] G. Mehl, Ö. Tunçalp, N. Ratanaprayul, T. Tamrat, M. Barreix, D. Lowrance, K. Bartolomeos, L. Say, N. Kostanjsek, R. Jakob, *et al.*, "Who smart guidelines: optimising country-level use of guideline recommendations in the digital age," *The Lancet Digital Health*, vol. 3, no. 4, pp. e213–e216, 2021.

[35] European Commision, "European Health Data Space." https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en, 2024.

[36] B. Blobel, P. Ruotsalainen, F. Oemig, M. Giacomini, P. A. Sottile, and F. Endsleff, "Principles and standards for designing and managing integrable and interoperable transformed health ecosystems," *Journal of Personalized Medicine*, vol. 13, no. 11, p. 1579, 2023.

[37] EU Observatory for ICT standardisation, "Health Informatics - Interoperability and integration reference architecture - Model and framework (ISO 23903:2021)." https://standict.eu/standards-repository/health-informatics-interoperability-and-integration-reference-architecture, 2024. Accessed: 2024-12-22.

[38] ISO, "ISO 23903:2021. Health informatics — Interoperability and integration reference architecture — Model and framework." https://www.iso.org/standard/77337.html, 2021. Accessed: 2024-12-22.

[39] E. Adel, S. El-Sappagh, S. Barakat, and M. Elmogy, "Ontology-based electronic health record semantic interoperability: A survey," in *U-healthcare monitoring systems*, pp. 315–352, Elsevier, 2019.

[40] M. Lehne, S. Luijten, P. Vom Felde Genannt Imbusch, and S. Thun, "The use of fhir in digital health–a review of the scientific literature," *German Medical Data Sciences: Shaping Change–Creative Solutions for Innovative Medicine*, pp. 52–58, 2019.

[41] K. H. Hafseld, B. Hussein, and A. R. Rauzy, "Government inter-organizational, digital transformation projects: five key lessons learned from a Norwegian case study," *Procedia Computer Science*, vol. 196, pp. 910–919, 2022.

[42] R. Kavanagh and N. I. C. Domain, "How is nhs digital using fhir," 2017.

[43] A. Metke-Jimenez, J. Steel, D. Hansen, and M. Lawley, "Ontoserver: a syndicated terminology server," *Journal of biomedical semantics*, vol. 9, no. 1, pp. 1–10, 2018.

[44] G. Lichtner, B. S. Alper, C. Jurth, C. Spies, M. Boeker, J. J. Meerpohl, and F. von Dincklage, "Representation of evidence-based clinical practice guideline recommendations on FHIR," *Journal of Biomedical Informatics*, vol. 139, p. 104305, 2023.

[45] HL7, "FHIR Shorthand." https://hl7.org/fhir/uv/shorthand, 2024. Accessed: 2024-12-22.

[46] N. Hong, K. Wang, S. Wu, F. Shen, L. Yao, and G. Jiang, "An interactive visualization tool for HL7 FHIR specification browsing and profiling," *Journal of healthcare informatics research*, vol. 3, pp. 329–344, 2019.

[47] P. L. Elkin and S. H. Brown, "Unified Medical Language System (UMLS)," in *Terminology, Ontology and their Implementations*, pp. 463–474, Springer, 2023.

[48] G. Kleinoscheg, N. Tanjga, N. Svec, S. Rainer-Sablatnig, and S. Sabutsch, "TerminoloGit-An Open Source Terminology Server for Large Scale eHealth Environments.," in *dHealth*, pp. 79–84, 2022.

[49] Firely, "Simplifier." https://simplifier.net, 2024. Accessed: 2024-12-22.

[50] J. Metsallik, P. Ross, D. Draheim, and G. Piho, "Ten years of the e-health system in estonia," in *CEUR Workshop Proceedings* (A. Rutle, Y. Lamo, W. MacCaull, and L. Iovino, eds.), vol. 2336, pp. 6–15, 3rd International Workshop on (Meta)Modelling for Healthcare Systems (MMHS), 2018.

[51] I. Bossenko, G. Piho, R. Bondarev, and P. Ross, "Visual FML Editor for data transformations by analysts using TermX," *Zenodo*, 2025. https://zenodo.org/records/14608721.

[52] F. Cascini, A. Pantovic, Y. A. Al-Ajlouni, V. Puleo, L. De Maio, and W. Ricciardi, "Health data sharing attitudes towards primary and secondary use of data: a systematic review," *EClinicalMedicine*, vol. 71, 2024.

[53] R. Õitspuu, "An Analysis of the Data Composition of the Estonian National Health Information System and Recommendations for the Implementation of the HL7 FHIR standard," Master's thesis, Tallinn University of Technology, 2024. https://digikogu.taltech.ee/et/Item/6b137f0a-efda-42b1-9653-d593d317f712. Accessed: 2024-12-22.

[54] ISO, "ISO 18308:2011. Health informatics — Requirements for an electronic health record architecture.." https://www.iso.org/standard/52823.html, 2023. Accessed: 2024-12-22.

[55] A. Hevner, S. Chatterjee, A. Hevner, and S. Chatterjee, "Design science research in information systems," *Design research in information systems: theory and practice*, pp. 9–22, 2010.

[56] M. Indulska and J. Recker, "13. design science in is research: a literature analysis," *Information systems foundations: The role of design science*, pp. 285–302, 2010.

[57] R. J. Wieringa, *Design Science methodology for information systems and software engineering*. Springer, 2014.

[58] HL7, "FHIR Specification." http://fhir.hl7.org, 2024. Accessed: 2024-12-22.

[59] HL7, "HL7 Common Terminology Services Service Functional Model Specification." http://www.hl7.org/documentcenter/private/standards/CTS/V3_CTS2_r2_2015FEB_R2022.pdf, 2022. Accessed: 2024-12-22.

[60] M. Ivanova, I. Bossenko, and G. Piho, "TermX — Bridging the Gap: Implementing CTS2 and FHIR Compatible Terminology Server," in *Proceedings of Ninth International Congress on Information and Communication Technology*, vol. 1011 of *Lecture Notes in Networks and Systems*, pp. 277–287, Springer Nature Singapore, Springer, 2024.

[61] I. Bossenko, J. Metsallik, G. Piho, and P. Ross, "Terminology quality Checklist for medical terminology developers and evaluators," *Zenodo*, 2025. https://zenodo.org/records/14609494.

[62] HL7, "FHIR StructureDefinition." http://hl7.org/fhir/structuredefinition.html, 2024. Accessed: 2024-12-22.

[63] HL7, "FHIR Mapping Language." http://hl7.org/fhir/mapping-language.html, 2024. Accessed: 2024-12-22.

[64] HL7, "HL7 CIMI Logical Model: Analysis Normal Form (ANF), Release 1." https://confluence.hl7.org/pages/viewpage.action?pageId=66925195, 2020.

[65] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo, "HL7 clinical document architecture, release 2," *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 30–39, 2006.

[66] Markdown-it community, "Markdown-it processor." https://github.com/markdown-it/markdown-it, 2024. Accessed: 2024-12-22.

[67] Mitre, "FSH documentation." https://fshschool.org/docs, 2024. Accessed: 2024-12-22.

[68] HL7, "FHIR StructureDefinition." http://hl7.org/fhir/structuredefinition.html, 2024. Accessed: 2024-12-22.

[69] A. Dimitrov and G. Duftschmid, "Generation of FHIR-Based International Patient Summaries from ELGA Data.," in *dHealth 2022*, vol. 293 of *Studies in Health Technology and Informatics*, pp. 1–8, Spinger, 2022.

[70] S. Mohan, I. F. Adams, S. K. Nallamalla, and A. Sharma, "Simplified near data placement processing layout for distributed storage object systems," in *2022 IEEE Women in Technology Conference (WINTECHCON)*, pp. 1–6, IEEE, 2022.

[71] A. M. Rajput, K. Triep, and O. Endrich, "Semi-Automated Approach to Map Clinical Concepts to SNOMED CT Terms by Using Terminology Server.," *Studies in health technology and informatics*, vol. 293, pp. 67–72, 2022.

[72] D. Divyabharathi and N. G. Cholli, "A review on identity and access management server (keycloak)," *International Journal of Security and Privacy in Pervasive Computing (IJSPPC)*, vol. 12, no. 3, pp. 46–53, 2020.

[73] M. Trifan, B. Ionescu, and D. Ionescu, "A Combined Finite State Machine and PlantUML Approach to Machine Learning Applications," in *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 000631–000636, IEEE, 2023.

[74] TermX, "TermX GitHub repository." https://github.com/termx-health, 2024. Accessed: 2024-12-22.

[75] TermX, "TermX download." https://termx.org/download, 2024. Accessed: 2024-12-22.

[76] TermX, "TermX quick start." https://termx.org/quick-start, 2024. Accessed: 2024-12-22.

[77] TermX, "TermX homepage." https://termx.org, 2024. Accessed: 2024-12-22.

[78] TermX, "TermX tutorial." https://tutorial.termx.org/en/fml-transformations, 2024. Accessed: 2024-12-22.

[79] M. T. Cabré, *Terminology: Theory, methods, and applications*, vol. 1. John Benjamins Publishing, 1999.

[80] ISO, "ISO 1087:2019. Terminology work and terminology science — Vocabulary.." https://www.iso.org/obp/ui/en/#iso:std:iso:1087:ed-2:v1:en, 2019. Accessed: 2024-12-22.

[81] B. Smith, W. Ceusters, and R. Temmerman, "Wüsteria," *Studies in Health Technology and Informatics*, pp. 647–652, 2005.

[82] J. J. Cimino, "Desiderata for controlled medical vocabularies in the twenty-first century," *Methods of information in medicine*, vol. 37, no. 04/05, pp. 394–403, 1998.

[83] B. Smith, "New desiderata for biomedical terminologies," *Applied Ontology*, pp. 83–109, 2008.

[84] C. J. Koné, M. Babri, and J. M. Rodrigues, "SNOMED CT: A Clinical Terminology but Also a Formal Ontology," *Journal of Biosciences and Medicines*, vol. 11, no. 11, pp. 326–333, 2023.

[85] L. Amos, D. Anderson, S. Brody, A. Ripple, and B. L. Humphreys, "UMLS users and uses: a current overview," *Journal of the American Medical Informatics Association*, vol. 27, no. 10, pp. 1606–1611, 2020.

[86] P. Zanstra, A. Rector, W. D. Solomon, T. Rush, W. Nowlan, S. Bechhofer, and N. Netherlands, "A terminology server for integrating clinical information systems: the galen approach," *CURRENT PERSPECTIVES IN HEALTHCARE COMPUTING*, pp. 256–261, 1995.

[87] R. Cornet and A. K. Prins, "An architecture for standardized terminology services by wrapping and integration of existing applications," in *AMIA Annual Symposium Proceedings*, vol. 2003, p. 180, American Medical Informatics Association, 2003.

[88] E. Aranda, J. L. V. Noguera, and C. Villalba, "Medical terminology server for the hospital of clinics of paraguay using apache lucene and the umls metathesaurus," *CLEI ELECTRONIC JOURNAL*, vol. 22, no. 3, 2019.

[89] E. F. Codd, *The relational model for database management: version 2*. Addison-Wesley Longman Publishing Co., Inc., 1990.

[90] OMG, "The Object Management Group® Standards Development Organization (OMG® SDO)." https://www.omg.org, 2024. Accessed: 2024-12-22.

[91] OMG, "Model Driven Architecture® (MDA®)." https://www.omg.org/mda, 2024. Accessed: 2024-12-22.

[92] A. W. Brown, "Model driven architecture: Principles and practice," *Software and systems modeling*, vol. 3, pp. 314–327, 2004.

[93] M. A. Kramer, "Fhir "profiliferation": A data science approach," *medRxiv*, 2022.

[94] E. H. Sibley and R. W. Taylor, "A data definition and mapping language," *Communications of the ACM*, vol. 16, no. 12, pp. 750–759, 1973.

[95] H. H. Rodriguez and D. S. Kolovos, "Declarative Model Transformation Execution Planning.," in *OCL@ MoDELS*, vol. 1756, pp. 105–120, CEUR-WS, 2016.

[96] I. Kurtev, "State of the art of QVT: A model transformation language standard," in *Applications of Graph Transformations with Industrial Relevance: Third International Symposium, AGTIVE 2007, Kassel, Germany, October 10-12, 2007, Revised Selected and Invited Papers 3*, (Kassel,Germany), pp. 377–393, Springer, 2008.

[97] Object Management Group (OMG), "Meta Object Facility (MOF) 2.0 Query/View/-Transformation Specification. Version 1.3." https://www.omg.org/spec/QVT/About-QVT, 2016. Accessed: 2024-12-22.

[98] A. Kras, W. Oliver, M. C. Gillies, M. Mair, S. Chu, N. Morlet, M. Mbagwu, T. S. Hwang, M. Roe, and T. Leng, "Extending interoperability in ophthalmology through fast healthcare interoperability resource (fhir)," *Investigative Ophthalmology & Visual Science*, vol. 62, no. 8, pp. 1708–1708, 2021.

[99] C. Erbelding, B. Sailer, H. Stenzhorn, S. Biergans, F. Kohlmayer, E.-M. Kobak, A.-A. Pape, R. Verbücheln, and O. Kohlbacher, "GECCO on FHIR-towards interoperable data on COVID-19," in *Proceedings of the 66th Annual Conference of the German Association for Medical Informatics*, pp. 26–30, 2021.

[100] HL7, "FHIR Open Source Implementations." https://confluence.hl7.org/display/FHIR/Open+Source+Implementations, 2024. Accessed: 2024-12-22.

[101] I. Bossenko, K. Linna, G. Piho, and P. Ross, "Migration from HL7 CDA to FHIR in Infectious Disease System of Estonia," in *pHealth 2022*, Studies in Health Technology and Informatics, pp. 275–278, IOS Press, 2022.

[102] M. Marquis, I. Bossenko, and P. Ross, "Standardizing X-Ray and Angiography Procedure Coding in Estonia: Validating TermX Through the Implementation of SNOMED CT," in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, vol. 316 of *Studies in Health Technology and Informatics*, pp. 53–54, IOS Press, 2024.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Gunnar Piho and Peeter Ross, for their unwavering support, insightful guidance, and invaluable feedback throughout this journey.

To Gunnar Piho, your mentorship and dedication have been a constant source of inspiration.

This thesis would not have been possible without the support of my family.

To my wife, Ljudmila Linnik, your constant support and encouragement have been my anchor.

This work in the 'ICT programme' project was supported by the European Union through the European Social Fund.

Abstract

A domain-specific framework for supporting semantic interoperability in primary and secondary use of health data on the example of the Estonian National Health Information System

This thesis investigates the semantic interoperability in the primary and secondary use of health data. Targeting both semantic and syntactic interoperability, the research addresses critical challenges in healthcare data sharing, particularly within the context of the Refined eHealth European Interoperability Framework (ReEIF). The study develops and evaluates TermX, a platform designed to improve interoperability in healthcare information systems. TermX includes terminology servers, data models, transformation editors, and publishing tools aimed at enabling the seamless exchange and accurate interpretation of electronic health records (EHRs) across healthcare providers and borders.

Motivation. The increasing digitalisation of healthcare necessitates interoperable EHR systems that facilitate seamless data exchange across providers and national boundaries. Despite various standards, the complexity of healthcare data management and the lack of common solutions continue to hinder semantic and syntactic interoperability. This thesis is motivated by the need to simplify the healthcare data sharing process, reduce implementation costs, and provide accessible, user-friendly solutions for healthcare interoperability, specifically targeting non-technical users who play essential roles in data management.

Problem definition. The central issue addressed by this research is the absence of a user-friendly, cohesive platform for managing healthcare data interoperability that effectively bridges technical and semantic gaps in data exchange. Current solutions often operate in isolation, require significant technical knowledge, or fail to fully support healthcare standards, thereby limiting their effectiveness in real-world applications. This thesis aims to create a platform that facilitates interoperability by unifying terminology management, data modelling, and transformation processes.

Methodology. The thesis employs a design science (DS) approach, utilising iterative cycles of problem investigation, artefact design, validation, and implementation. The initial cycle focused on justifying and constructing TermX, while subsequent cycles assessed its suitability for real-world scenarios and its implementation within national health services.

Novelty. This research introduces TermX, an open-source platform that addresses existing interoperability challenges with a modular approach integrating a terminology server, a data model designer, a transformation editor, and publication tools. Unlike previous frameworks, TermX is designed to support multilingual environments, facilitate complex data transformations (e.g. HL7 CDA to FHIR), and enable seamless integration with widely used standards. Its innovative aspect lies in its adaptability to various interoperability standards and its focus on usability, making it accessible to analysts and healthcare workers without extensive technical expertise.

Results. TermX has demonstrated effective results in multiple real-world use cases, facilitating accurate and efficient data standardisation methods among diverse healthcare entities. The platform successfully reduced the learning curve for non-technical users, streamlined the development of interoperable data models, and improved the management of healthcare terminology. The application of TermX in Estonian and Lithuanian healthcare systems underscores its utility and potential for broader adoption.

Conclusions. This research contributes to the field of healthcare interoperability by providing a scalable, adaptable, and accessible platform that addresses both syntactic and

semantic interoperability challenges. TermX offers an advanced solution for healthcare data management, reducing the need for technical expertise while ensuring compliance with international standards. The platform's successful deployment in various countries highlights its effectiveness and potential for future integration into broader healthcare systems. Future research will explore the expansion of TermX's capabilities and further refine its user interface and interoperability with additional standards.

Kokkuvõte

Tervishoiu domeenipõhise semantilise raamistiku loomine terviseandmete esmaseks ja teiseseks kasutamiseks Eesti Tervise infosüsteemi näitel

Antud uurimistöö uurib semantilist koostalitlusvõimet terviseandmete esmase ja teisese kasutamise kontekstis. Keskendudes nii semantilisele kui ka süntaktilisele koostalitlusvõimele, käsitleb uurimus kriitilisi väljakutseid tervishoiu andmete jagamisel, eriti Euroopa e-tervise täiustatud koostalitlusvõime raamistiku (ReEIF) kontekstis. Uurimistöö raames arendatakse ja valideeritakse TermX-i, platvormi, mis on loodud tervishoiu infosüsteemide koostalitlusvõime parandamiseks. TermX sisaldab terminoloogia serverit, andmemudelite modeleerimise, transformatsioonide haldamise ning publitseerimise tööriistu, mille eesmärk on võimaldada elektrooniliste terviseandmete (EHR) sujuvat vahetust ja täpset tõlgendamist tervishoiuteenuste osutajate vahel.

Motivatsioon. Tervishoiu digitaliseerimise kasv nõuab koostalitlusvõimelisi elektroonilisi terviseandmete (EHR) süsteeme, mis hõlbustavad sujuvat andmevahetust teenuseosutajate vahel ka riigipiiride üleselt. Vaatamata erinevatele standarditele takistavad tervishoiu andmete haldamise keerukus ja ühiste lahenduste puudumine endiselt semantilist ja süntaktilist koostalitlusvõimet. Antud uurimistöö on ajendatud vajadusest muuta tervishoiu andmete jagamist lihtsamaks, vähendada kulusid ja pakkuda ligipääsetavaid ning kasutajasõbralikke lahendusi tervishoiu koostalitlusvõime jaoks. Töö sihtrühmana on mittetehnilised kasutajad, kes mängivad olulist rolli andmete haldamisel.

Probleemi määratlus. Selle uurimistöö peamine uurimisprobleem on kasutajasõbraliku ja ühtse platvormi puudumine, mis suudaks tõhusalt lahendada tehnilised ja semantilised lüngad tervishoiu andmete koostalitluses. Praegused lahendused toimivad sageli isoleeritult, nõuavad märkimisväärseid tehnilisi teadmisi või ei toeta täielikult tervishoiu standardeid, piirates seeläbi nende tõhusust reaalsetes rakendustes. See doktoritöö kavandab ja valideerib platvormi, mis hõlbustab koostalitlusvõimet, ühendades terminoloogia, andmemudelite ja transformatsioonide koostamist ja haldamist.

Metoodika. Uurimistöös kasutatakse disainiteaduse (DS) lähenemist, rakendades iteratiivseid tsükleid probleemi uurimiseks, artefakti kujundamiseks, valideerimiseks ja rakendamiseks. Esimene disainitsükkel keskendus TermX-i põhjendamisele ja ehitamisele, samas kui järgnevad tsüklid hindasid selle sobivust reaalsetes stsenaariumides ja selle rakendamist riiklike tervishoiuteenuste pakkumisel.

Uudsus. Antud uurimustöö tutvustab TermX-i, avatud lähtekoodiga platvormi, mis läheneb koostalitlusvõime probleemile modulaarselt. TermX ühendab terminoloogia serveri, andmemudeli kujundaja, teisendustoimetaja ja avaldamise tööriistad. Erinevalt varasematest platvormidest on TermX loodud toetama mitmekeelseid keskkondi, hõlbustama keerukaid andmete teisendusi (nt HL7 CDA-st FHIR-iks) ja võimaldama sujuvat integreerimist erinevate laialdaselt kasutatavate standarditega. TermX uuenduslikkus seisneb selle kohanemisvõimes erinevate koostalitlusvõime standarditega ja kasutatavuse lihtsuses, muutes selle kättesaadavaks analüütikutele ja tervishoiutöötajatele kellel tavaliselt pole pisavalt tehnilisi teadmisi.

Tulemused. TermX on näidanud tõhusaid tulemusi mitmes reaalses kasutusjuhtumis, võimaldades täpseid ja tõhusaid andmete koostalitlust erinevate tervishoiuüksuste vahel. Platvorm on edukalt vähendanud mitte-tehniliste kasutajate õppimiskõverat, lihtsustanud koostalitlusvõimeliste andmemudelite loomist ja parandanud tervishoiu terminoloogia haldamist. TermX-i rakendamine Eesti ja Leedu tervishoiusüsteemides näitab selle kasulikkust ja potentsiaali laiemaks kasutuselevõtuks.

Järeldused. Antud uurimustöö panustab tervishoiu koostalitlusvõime valdkonda, pakkudes skaleeritavat, kohandatavat ja ligipääsetavat platvormi, mis aitab lahendada semantilise koostalitlusvõime väljakutseid. TermX pakub täiustatud lahendust tervishoiu andmete haldamiseks, vähendades oluliselt tehniliste teadmiste vajadust, tagades samal ajal vastavuse rahvusvahelistele standarditele. Platvormi edukas kasutuselevõtt erinevates riikides näitab selle tõhusust ja potentsiaali tulevaseks laialdaseks integreerimiseks erinevatesse tervishoiusüsteemidesse. Tulevased võimalikud uurimustöö suunad keskenduvad TermX-i võimekuse laiendamisele, kasutajaliidese täiendamisele ning koostalitlusvõimele uute standarditega.

Appendix 1

I

Bossenko, Igor and Piho, Gunnar and Ivanova, Marina and Ross, Peeter, “TermX: The semantic interoperability, knowledge management and sharing platform,” *SoftwareX*, vol. 27, 2024

SoftwareX 27 (2024) 101839

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

TermX: The semantic interoperability, knowledge management and sharing platform

Igor Bossenko [a,*], Gunnar Piho [a], Marina Ivanova [a], Peeter Ross [b]

[a] Department of Software Science, Tallinn University of Technology, Akadeemia 15A, 12618 Tallinn, Estonia
[b] Department of Health Technologies, Tallinn University of Technology, Estonia

ARTICLE INFO

Keywords:
TermX
Terminology server
Fast Healthcare Interoperability Resources (FHIR)
Model designer
FHIR Mapping Language (FML) editor

ABSTRACT

TermX is an open-source knowledge management and sharing platform, including a terminology server, a Wiki, a model designer, a transformation editor, and tools for authoring and publishing. The core development goals of TermX were to enhance the semantic interoperability of software and systems, particularly within the healthcare sector, by improving access to terminology, simplifying the design of data models, optimising the efficiency of data transformations across various data models, and developing implementation guides through an intuitive web interface. TermX aims to guarantee open and standardised access to published terminology, data models and schemas, ensuring semantic interoperability following the FHIR standard or other standards or agreements.

Code metadata

Current code version	2.5
Permanent link to source code	https://github.com/ElsevierSoftwareX/SOFTX-D-24-00220
Legal Code Licence	MIT
Code versioning system used	Git
Software code languages, tools, and services used	Java, PostgreSQL, Javascript, Angular, HTML, CSS, Keycloak, Snowstorm, PlantUML
If available, link to developer documentation/manual	https://termx.org https://termx.org/docs
Support email for questions	support@termx.org

1. Motivation and significance

Knowledge standardisation and semantic interoperability are essential across all clinical systems. Interoperable healthcare software and systems are pivotal for enhancing patient safety, improving efficiency, driving innovation, and facilitating seamless data exchange across diverse healthcare systems, devices, and applications. These advancements significantly impact global health outcomes [1].

Semantic interoperability and having data in a common format are crucial for efficient data exchange and reducing duplication, delays, waste, and errors [2]. Fast Healthcare Interoperability Resources (FHIR) has rapidly become the most important health interoperability standard globally [3]. FHIR allows developers to develop standardised browser applications that enable the user to access clinical data from any healthcare system, regardless of the operating systems and devices that healthcare system uses [4]. FHIR can be implemented at a fraction of the price of existing alternatives and is well suited for use in mobile phone apps, cloud communications and electronic health records [3], running third-party applications without expensive custom integration [5], and managing and transforming health data in the era of the Internet of medical things (IoMT) and 5G [6].

Individuals or small teams cannot manage or operate the development and change of large terminologies and standards; they require supportive tools [7]. Many countries have specialised organisations or departments for standardising healthcare terminologies [8–10]. Their tasks include: (1) developing terminology; (2) developing logical data

* Corresponding author.
E-mail addresses: igor.bossenko@taltech.ee (Igor Bossenko), gunnar.piho@taltech.ee (Gunnar Piho), marina.ivanova@taltech.ee (Marina Ivanova), peeter.ross@taltech.ee (Peeter Ross).

https://doi.org/10.1016/j.softx.2024.101839
Received 6 May 2024; Received in revised form 2 July 2024; Accepted 26 July 2024
Available online 8 August 2024

models to represent clinical information; (3) adapting interoperability standards such as FHIR [3] or HL7 Clinical Document Architecture (CDA) [11]; (4) designing data transformation between logical data models and interoperability standards; (5) creating a knowledge base and thesaurus; and (6) publishing standards and implementation guidelines.

Despite the potential of FHIR, research suggests that it encounters several obstacles, including its implementation, acceptance, upkeep, mapping, and the complexity of the standard [4].

The successful development and implementation of data models and terminology and the facilitation of interoperability necessitate suitable software. Numerous programs exist to perform specific tasks, such as Ontoserver for terminology [12], SUSHI for logical models [13,14], Forge for FHIR profiling [15], FHIR validator for data transformations [13], UMLS for thesaurus [16], and IG Publisher [13,17] and Simplifier [18] for Implementation Guide and publishing.

Many software programs demand extensive human expertise. Frequently, these applications operate in isolation, lacking integration within a cohesive ecosystem. Achieving interoperability among them poses a substantial challenge. A comprehensive assessment of existing terminology servers has revealed a deficiency [19]. There is a shortage of open-source, no-cost, high-quality tools that provide multilingual support and feature an advanced graphical user interface [19].

We introduce a series of papers that describe **TermX** as a platform for semantic interoperability, knowledge management and data sharing. This paper presents an overview of the platform. In subsequent publications, we plan to consider each platform module in detail.

The rest of the paper is organised as follows: Section 2 provides a comprehensive review of the functionality and architecture of TermX. Section 3 illustrates the examples of TermX use. Section 4 evaluates TermX and discusses the social impacts of its implementation. Finally, in Section 5, we summarise the results and outline directions for further research and publications.

2. Software description

TermX is a unique open-source interoperability platform encompassing various knowledge management and sharing components. It contains modules with unique capabilities, such as a visual designer of FHIR Mapping Language (FML) transformations, a terminology server supporting multilingual clinical terminology, multilingual resource descriptions, and a multilingual web user interface [19]. Furthermore, its internal data models are based on widely accepted standards, such as Common Terminology Services 2 (CTS2) [20] and FHIR, ensuring seamless integration and compatibility [21]. The main purpose of TermX development was to support specialists in developing implementation guidelines through an intuitive web user interface, including terminology, data models, transformations and publishing options. Although TermX can transform any data format to any other data format, this paper emphasises FHIR transformations.

2.1. Software architecture

TermX consists of the internally and externally developed microservices required to run TermX. The TermX micro-services include (Fig. 1): (1) **TermX server**, the main back-end application, written in Java, responsible for orchestrating TermX modules; (2) **TermX web**, an Angular-based front-end application, communicates with the server via a RESTful API, offers a user-friendly interface for effortless task execution; (3) **Chef**, the wrapper for Mitre Sushi and GoFSH projects [22], provides the ability to transform the FHIR JSON notation of the resources into Shorthand FSH notation [14] and vice versa. TermX web includes **FML editor** and FHIR **StructureDefinition viewer**, which are Javascript and Angular-based libraries for data transformation management. These libraries interact with StructureMap [23] resources while implementing the FHIR FML [24,25] specification.

The TermX server is composed of independent modules. Communication between these modules is organised through the 'termx-api' module, which provides communication interfaces. The implementation of the modules varies based on the tasks they perform, and they communicate with external micro-services as needed. The persistence layer utilises the relational and non-SQL features of the PostgreSQL database server. The terminology module, which employs the CTS2 specification, describes a variety of relatively fine-grained resources and associations. These are implemented as an Entity-Relationship model in the database [21].

TermX can be customised and extended to suit customer-specific needs and preferences. Most of the modules can be replaced with customised implementation. For example, implementing an internal task management module can be replaced with a Jira [26] adaptor. It is also possible to add new features, functionalities, and CSS styles to TermX by developing customised plugins or modules.

The external micro-services include (Fig. 1): (1) the **PostgreSQL database**, which is used as internal data storage in TermX; (2) a Minio-based [27] **binary object repository** for storing attachments and large files; (3) **GitHub** for synchronising terminology resources and Wiki pages and automating processes related to the FHIR Implementation Guide and static site generation; (4) third-party **FHIR-compatible terminology servers** that may be used as sources or destinations for the synchronisation of terminology resources; (5) the IHTSDO **Snowstorm** server and SCT browser [28] used for management of the SNOMED terminology; (6) the Keycloak **OpenID Connect server** used for user and role management [29]; and (7) the **PlantUML server** used to generate images of UML diagrams [30]. TermX components are available as source code, libraries, and Docker images [31]. The deployment configurations for Docker Compose and Helm simplify the installation process [32].

2.2. Software functionalities

2.2.1. Terminology server

The terminology server is based on CTS2 [20], a standard for accessing and managing terminologies. TermX empowers users to create, store, update, and query terminologies from diverse clinical and administrative domains and multiple human languages. TermX supports the FHIR terminology module, including code systems, code system supplements, value sets, concept maps, and naming systems. TermX also supports the versioning, validation, and authoring of terminologies. The TermX SNOMED module enables the management and translation of SNOMED concepts.

2.2.2. Model designer

The Model Designer implements the FHIR StructureDefinition specification [33], which prescribes rules for logical data models, FHIR resources, and profiles as well as their elements and data types. With Model Designer, creating and managing data models through a user-friendly interface is possible. For FHIR StructureDefinition, both JSON and FSH notations are supported. The model designer capabilities allow the import/export of models and define attributes, data types, cardinality, and terminology binding, among other features.

2.2.3. Visual editor for FML data transformations

The FHIR Mapping Language (FML) is a transformation language specifically designed to transform HL7 FHIR resources to/from alternative representations [25], including different logical data models [34], FHIR resources, CDA documents [35], etc. TermX provides a unique visual FML Editor as a designer of explicitly designed FML transformations for business analysts (Fig. 6). The objective of the FML editor is to visually represent transformations, hide the complexity of the FML language, and facilitate rapid adaptation to it.

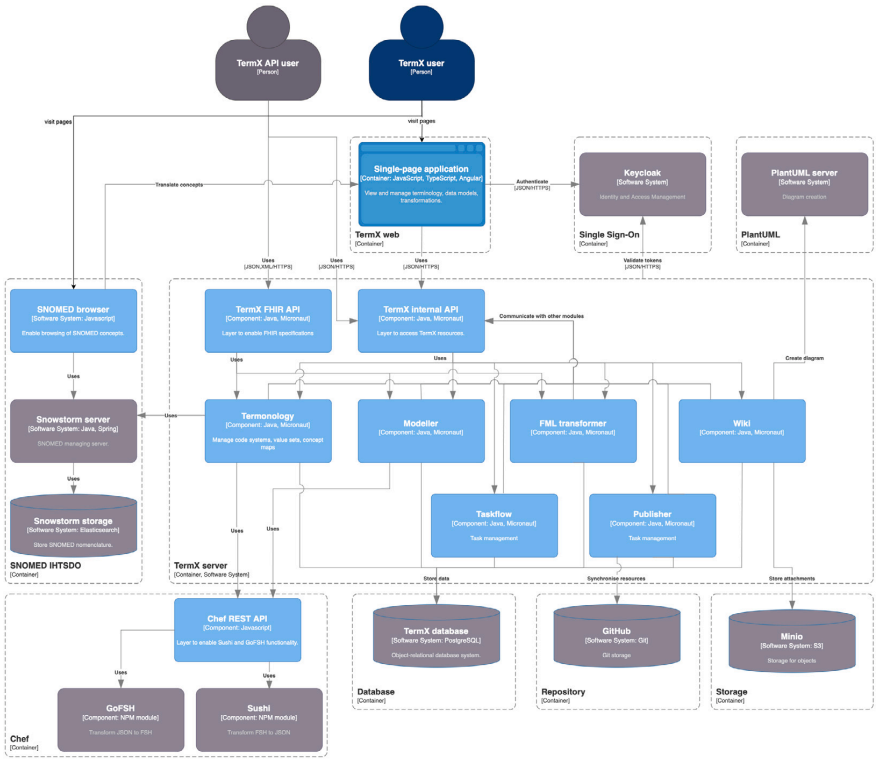

Fig. 1. TermX component diagram.

2.2.4. Wiki

A wiki is a web-based collaborative platform that enables well-formatted content creation and editing in an organised manner. It serves as a foundation for knowledge management, including thesaurus, tutorials, and custom pages for static sites and FHIR Implementation Guides. The thesaurus plays a crucial role in defining terms for effective communication. Additionally, Wiki plugins allow for easy referencing of terminology, models, and transformations and their inclusion on pages. Markdown syntax is commonly used to edit Wiki pages. Wiki implementation is based on Markdown-it processors [36].

2.2.5. Publisher

A publisher facilitates the organisation of terminology, models, and Wiki pages into logical spaces. It enables the bidirectional synchronisation of resources with external storage. Specifically, the publisher supports synchronisation with GitHub repositories and syndication with terminology servers that adhere to the FHIR Terminology API. Additionally, it provides the capability to generate Implementation Guides and static sites.

2.2.6. Task management

The task management system facilitates user collaboration to support the authoring process. Users can create tasks for reviewing and accepting new, modified, or translated concepts and versions of code systems, value sets, or concept maps in the terminology module. Additionally, they can create tasks to review Wiki pages or specific sections within those pages. Task management organises user work and provides an overview of unresolved tasks. This system ensures that all tasks are tracked and managed efficiently, streamlining workflow and enhancing productivity.

2.2.7. Chef

FSH (FHIR Shorthand) is a specially designed language for defining the content of HL7 FHIR Implementation Guides [22]. SUSHI (SUSHI Unshortens ShortHand Inputs) is an FSH compiler. SUSHI converts FSH language to FHIR artefacts. GoFSH is a converter that takes FHIR artefacts and produces equivalent FSH. GoFSH is essentially the opposite of SUSHI [22]. SUSHI and FSH implemented the Node Package Manager (NPM) libraries with many dependencies and specific syntax to execute. The Chef is the wrapper application for SUSHI and GoFSH. It provides a simple REST API to perform transformations between JSON and FSH and encapsulates them into a single Docker container.

3. Illustrative examples

In the following, we provide a limited set of examples to demonstrate some of the features of TermX. All screenshots and scripts in this section can be downloaded from the GitHub project [37].

3.1. The code system management in the terminology server

The web interface allows users to view and search for resources in the list (Fig. 2), navigate to a detailed view of a resource (Fig. 3), or add a new resource manually or through file importers from the FHIR, FSH, or CSV formats. Fig. 2 shows the importing of the ‘marital status’ FHIR code system from the HL7 Terminology site [38]. The terminology server supports versioning and concept management. The detailed view (Fig. 3) provides an overview of the main attributes of the code system, its versions, the history of changes (provenance), links with other artefacts, and the compliance of the code system with terminology rules (checklist) and allows users to go to the list of

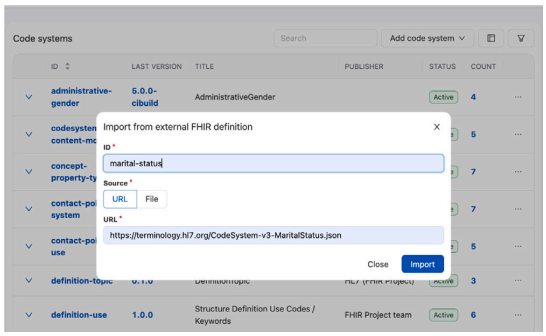

Fig. 2. Importing the code system.

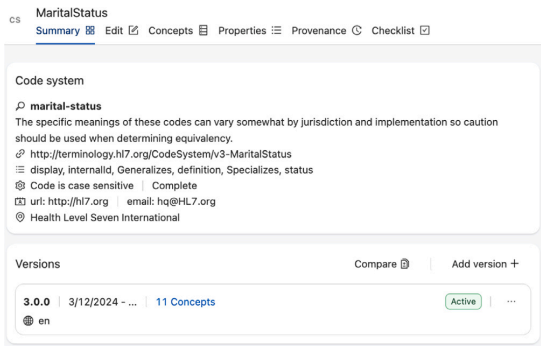

Fig. 3. Detailed view of the code system.

concepts and compare concepts between versions. The list of concepts (Fig. 4) can be represented as a hierarchy or a flat list that displays user-configurable properties. The concept preview feature allows users to get a complete picture of the code system without the need to navigate between screens.

3.2. Creating a logical model in the model designer

TermX allows the definition of logical models according to the HL7 FHIR StructureDefinition [33,39] standard. The model can be created with a visual designer or using FHIR JSON or FSH notation. Fig. 5 combines all three modes and demonstrates the logical model named SoftwareXPerson. SoftwareXPerson contains five attributes with different types and cardinalities. The black box in Fig. 5 shows how the same attribute 'code' is expressed in different modes. The Model Designer stores model definitions in the FHIR JSON format and transforms them between JSON and FSH using the Chef micro-service.

3.3. Transforming the logical model to the FHIR resource with a visual editor for FML data transformations

The TermX FML Editor is designed as a solution that generates an FHIR StructureMap from a visual representation. The visual FML Editor supports constants, functions, subtypes, concept maps, nested groups, and other elements defined in the FML specification. Fig. 6 visualises the data transformation from the SoftwareXPerson model described in the previous chapter to the FHIR Patient [40] resource. The value from 'SoftwareXPerson.birthdate' is copied to 'Patient.birthDate'. The 'uuid' function is used to set the 'Patient.id' value. The HumanName subtype is used to initialise 'Patient.name'. A constant is used to initialise the 'system' attribute of the subtype identification. Finally, ConceptMap maps the values of 'SoftwareXPerson.gender' to the values of the AdministrativeGender FHIR value set [40].

The TermX FML Editor, which utilises FML as the transformation language, relies on HAPI FHIR [41] as the foundation for the transformation engine and transforms data from input sources to output sources.

3.4. Wiki

A TermX Wiki allows users to organise knowledge into logical spaces. Fig. 7 illustrates the 'TermX tutorial' space, which is utilised for generating the TermX documentation website. Each space can support a hierarchical page structure, slugs, and multilingual content and store links to other pages, terminology, data models, and transformations. The wiki stores the history of pages and allows users to compare and restore pages. The wiki is an important component of the publisher.

4. Impact

The lack of interoperability between health information systems reduces patient care quality and wastes resources [42]. The semantic interoperability issue of medical data is a major challenge that healthcare systems face today. Data are often preserved in isolated databases, ensnared by incompatible systems, and constrained by proprietary software due to data privacy. These data prove challenging to exchange, analyse, and interpret effectively [43]. The terminology development assists interoperability and brings benefits through the consistent use of electronic health records [44].

TermX provides the platform for terminology management, simplifies data model design, and facilitates data transformations between models. It enhances interoperability, improves terminology consistency, publication and accessibility, fosters collaboration, and reduces maintenance costs.

TermX has changed its users' daily practices in several ways. It enables the precise definition and management of medical terms, ensuring the quality, consistency, and interoperability of terminologies and data across healthcare systems. The platform's intuitive web user interface and data exchange features with external services have significantly improved the efficiency of terminology and data model definition and management. Furthermore, the visual editor of data transformation hides the complexity of the FML language, allowing for wider adoption. In addition, TermX reduces the time required to learn the basics of terminology, optimises human resources, and simplifies the adaptation of the standards. Lastly, TermX promotes stakeholder collaboration through a shared understanding of medical terminology.

The adaptable and flexible TermX architecture allows integration into existing customer ecosystems. TermX has been developed and tested with TalTech, the national standardisation agencies of Estonia, Lithuania and Uzbekistan, and the private sector. There is an opportunity for the customer to receive support and implement and develop new features in collaboration with the TermX service provider. TermX has been incorporated into the healthcare interoperability and quality management course at TalTech University.

TermX was not designed to compete with existing software products. Instead, it aims to integrate with them and leverage their best features. For instance, the Forge Editor [45] currently exceeds TermX's capabilities in FHIR profiling. However, TermX enables the storage and utilisation of FHIR profiles created by Forge. Users tend to prefer TermX due to the relative simplicity of its web interface, the ability to customise this interface, and its integration with their internal software products. Other appealing features include enterprise Single Sign-On (SSO), the option of on-premise installation, and the absence of vendor lock-in, thanks to the open MIT licence.

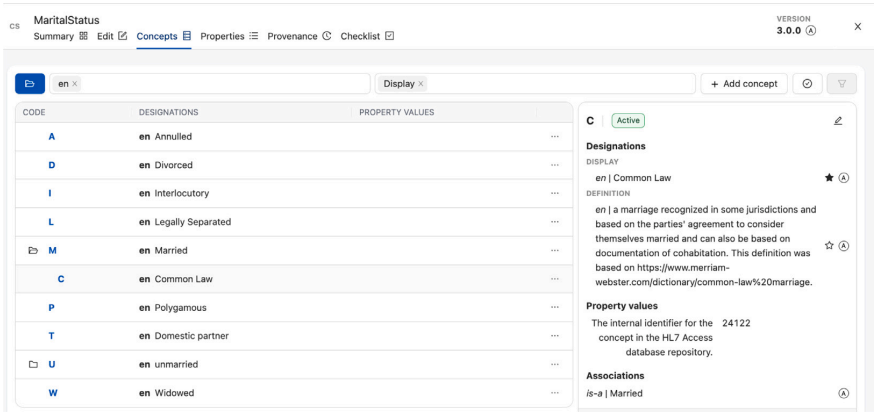

Fig. 4. The list of code system concepts and concept preview.

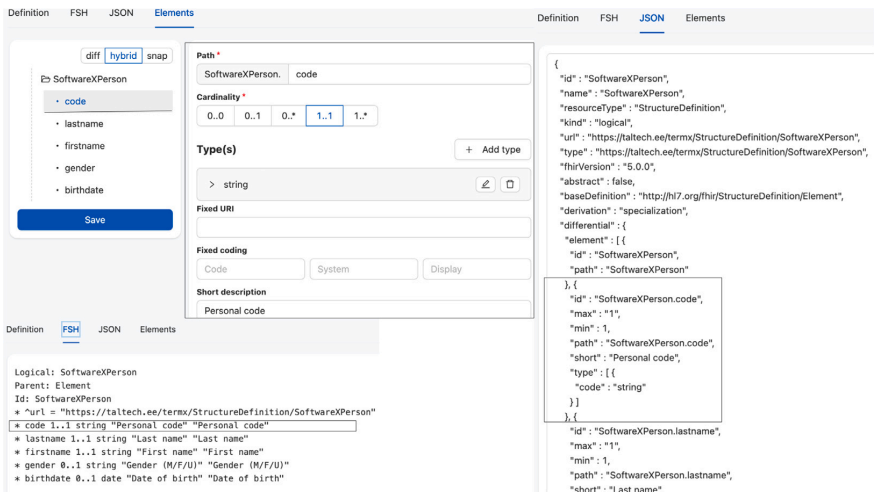

Fig. 5. Three presentations of the SoftwareXPerson model in Model Designer.

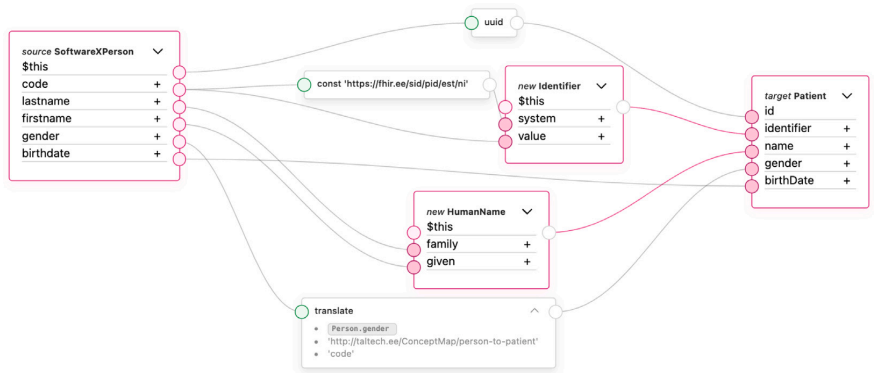

Fig. 6. Visualisation of the transformation from SoftwareXPerson model to FHIR Patient resource.

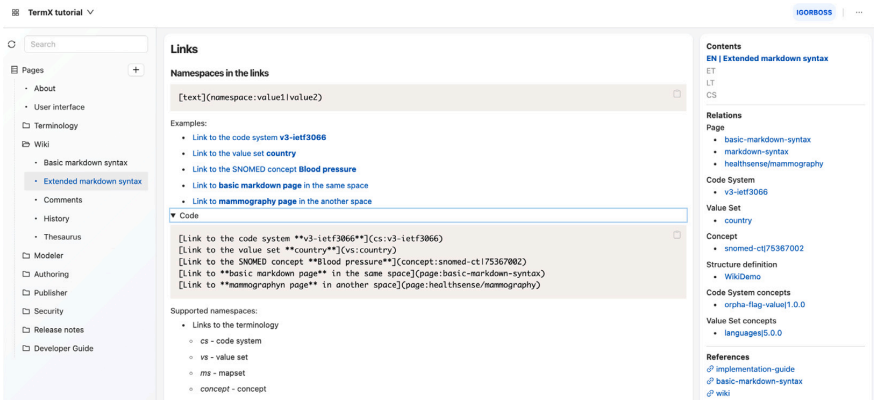

Fig. 7. Embedded content editor.

5. Conclusions

This paper describes TermX, the platform for semantic interoperability, and knowledge management and sharing. TermX provides a modern user interface for managing the terminology, data models, and transformations. TermX and FHIR API enable the automation of various tasks combined with other tools and scripts. Advanced publishing abilities such as synchronisation with GitHub, syndication with other terminology servers, and the generation of static sites and implementation guides enable information sharing over different channels.

The development of TermX was accompanied by a constant selection of the most suitable standards, technologies, and experts. While CTS2 is a stable standard that fully meets our terminology requirements, FHIR is still under continuous development. Answers to many questions regarding implementing the FHIR standard were received directly from the FHIR Director and other specialists via chat.fhir.org. Several topics have prompted modifications to the FHIR specification. Consultations with lead FML experts helped to create the FML Editor. The study of Ontoserver stimulated the possibility of syndication between servers. Implementing TermX in different countries, especially Estonia, helped to identify and solve various bugs and issues in TermX. The creation of such a product would not have been possible without the participation of international experts from various domains, including healthcare interoperability, implementation, academia, and technology.

We plan to cover each module in detail in subsequent scientific publications and compare them with existing tools. Development of TermX commenced in 2022. TermX is still a relatively new and developing product. We plan to enhance TermX's functionality and investigate the potential of alternative implementations. We aim to analyse and compare the types of databases used for CTS2-based terminology management, including NoSQL [46] and ontology [47] databases. Additionally, we intend to incorporate support for UML notation and FHIR profiling in the model designer, among other improvements.

The main goal of this research is to improve semantic interoperability and facilitate non-technical participation in the development of terminology, data models, and transformations. We expect this endeavour to significantly contribute to the semantic interoperability domain and its further progress.

CRediT authorship contribution statement

Igor Bossenko: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis. **Gunnar Piho:** Writing – review & editing, Supervision. **Marina Ivanova:** Writing – review & editing, Software. **Peeter Ross:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The software published as open-source. The paper contains links to the repository.

Acknowledgements

This work in the project 'ICT programme' was supported by the European Union through the European Social Fund. The EU Programme for Applied Research of Enterprise Estonia, TalTech, TEHIK, and Kodality supported the development of TermX. This software and its research would not have been possible without the exceptional support of Daniel Dubrovski, Roman Bondarev, Rutt Lindström, Jose Costa Teixeira, Oliver Egger, and Grahame Grieve.

References

[1] ISO. Sector: Health. 2024, https://www.iso.org/sectors/health. [Accessed: 01 July 2024].
[2] Kiourtis A, Nifakos S, Mavrogiorgou A, Kyriazis D. Aggregating the syntactic and semantic similarity of healthcare data towards their transformation to HL7 FHIR through ontology matching. Int J Med Informat 2019;132:104002.
[3] Benson T, Grieve G. Principles of health interoperability. Cham: Springer International; 2021, p. 21–40.
[4] Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The fast health interoperability resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Informat 2021;9(7):e21929.
[5] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Informat Assoc 2016;23(5):899–908.
[6] Mavrogiorgou A, Kiourtis A, Touloupou M, Kapassa E, Kyriazis D. Internet of medical things (IoMT): acquiring and transforming data into HL7 FHIR through 5G network slicing. Emerg Sci J 2019;3(2):64–77.
[7] Schmitz K-D, Straub D. Successful terminology management in companies. 2010, p. 772–87, Stuttgart. TC and more GmbH. Shreve, Gregory (2001). Terminological Aspects of Text Production. Handbook of Terminology Management, Vol. 2.
[8] Pirnejad H, Bal R, Stoop A, Berg M. Inter-organisational communication networks in healthcare: centralised versus decentralised approaches. Int J Integr Care 2007.
[9] Schliemann T, Danielsen C, Virtanen T, Vuokko R, Hardardottir GA, Alsaker MA, et al. eHealth standardisation in the Nordic countries:: Technical and partially semantics standardisation as a strategic means for realising national policies in eHealth. Nordic Council of Ministers; 2019.

[10] Egelkraut R, Traxler B, Helm E, Krauss O, Schuler A. Open infrastructure for standardization of HL7® FHIR® implementation guides in Austria. In: DHealth. 2022, p. 221–3.
[11] Mercorella M, Ciampi M, Esposito M, Esposito A, De Pietro G. An architectural model for extracting FHIR resources from CDA documents. In: 2016 12th International conference on signal-image technology & internet-based systems. IEEE; 2016, p. 597–603.
[12] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. J Biomed Semantics 2018;9:1–10.
[13] Lichtner G, Alper BS, Jurth C, Spies C, Boeker M, Meerpohl JJ, et al. Representation of evidence-based clinical practice guideline recommendations on FHIR. J Biomed Inform 2023;139:104305.
[14] HL7. FHIR shorthand. 2024, https://hl7.org/fhir/uv/shorthand/. [Accessed: 01 Jyly 2024].
[15] Hong N, Wang K, Wu S, Shen F, Yao L, Jiang G. An interactive visualization tool for HL7 FHIR specification browsing and profiling. J Healthc Informat Res 2019;3:329–44.
[16] Elkin PL, Brown SH. Unified medical language system (UMLS). In: Terminology, ontology and their implementations. Springer; 2023, p. 463–74.
[17] Kleinoscheg G, Tanjga N, Svec N, Rainer-Sablatnig S, Sabutsch S. TerminoloGit-an open source terminology server for large scale ehealth environments. In: DHealth. 2022, p. 79–84.
[18] Firely. Simplifier. 2024, https://simplifier.net. [Accessed: 01 Jyly 2024].
[19] Ivanova M, Bossenko I, Piho G. The comparison of the terminology servers. Springer; 2024, p. 12, is accepted by the 9th International Conference on Digital Economy (ICDEc), Rabat, Morocco.
[20] HL7. HL7 common terminology services. Service functional model specification. Release 2. 2024, http://www.hl7.org/documentcenter/private/standards/CTS/V3_CTS2_r2_2015FEB_R2022.pdf. [Accessed: 01 Jyly 2024].
[21] Ivanova M, Bossenko I, Piho G. TermX - Bridging the gap: Implementing CTS2 and FHIR compatible terminology server. In: International congress on information and communication technology. Springer; 2024, p. 10, is accepted by ICICT 2024 (9th International Congress on Information and Communication Technology), London, UK.
[22] Mitre. FSH documentation. 2024, https://fshschool.org/docs. [Accessed: 01 Jyly 2024].
[23] HL7. FHIR StructureMap. 2024, https://hl7.org/fhir/structuremap.html. [Accessed: 01 Jyly 2024].
[24] Dimitrov A, Duftschmid G. Generation of FHIR-based international patient summaries from elga data. In: DHealth. 2022, p. 1–8.
[25] HL7. FHIR mapping language. 2024, https://hl7.org/fhir/mapping-language.html. [Accessed: 01 Jyly 2024].
[26] Attlassian. JIRA. 2024, https://www.atlassian.com/software/jira. [Accessed: 01 Jyly 2024].
[27] Mohan S, Adams IF, Nallamalla SK, Sharma A. Simplified near data placement processing layout for distributed storage object systems. In: 2022 IEEE women in technology conference. IEEE; 2022, p. 1–6.
[28] Rajput AM, Triep K, Endrich O. Semi-automated approach to map clinical concepts to SNOMED CT terms by using terminology server. Stud Health Technol Informat 2022;293:67–72.
[29] Divyabharathi D, Cholli NG. A review on identity and access management server (keycloak). Int J Secur Privacy Pervasive Comput (IJSPPC) 2020;12(3):46–53.
[30] Trifan M, Ionescu B, Ionescu D. A combined finite state machine and plantuml approach to machine learning applications. In: 2023 IEEE 17th international symposium on applied computational intelligence and informatics. IEEE; 2023, p. 000631–6.
[31] TermX. TermX download. 2024, https://termx.org/download. [Accessed: 01 Jyly 2024].
[32] TermX. TermX quick start. 2024, https://termx.org/quick-start. [Accessed: 01 Jyly 2024].
[33] HL7. FHIR StructureDefinition. 2024, https://hl7.org/fhir/structuredefinition.html. [Accessed: 01 Jyly 2024].
[34] HL7. HL7 CIMI logical model: Analysis normal form (ANF), release 1. 2024, https://confluence.hl7.org/pages/viewpage.action?pageId=66925195. [Accessed: 01 Jyly 2024].
[35] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 clinical document architecture, release 2. J Am Med Informat Assoc 2006;13(1):30–9.
[36] Markdown-it community. Markdown-it processor. 2024, https://github.com/markdown-it/markdown-it. [Accessed: 01 Jyly 2024].
[37] TermX. Images and examples applied to the current paper. 2024, https://github.com/termx-health/knowledge-base/tree/main/softwarex-2024-termx-platform. [Accessed: 01 Jyly 2024].
[38] HL7. HL7 terminology. 2024, https://terminology.hl7.org. [Accessed: 01 Jyly 2024].
[39] Braunstein ML. FHIR. In: Health informatics on FHIR: how HL7's API is transforming healthcare. Springer; 2022, p. 233–91.
[40] HL7. FHIR patient. 2024, http://hl7.org/fhir/patient.html. [Accessed: 01 Jyly 2024].
[41] Smile Digital Health. HAPI FHIR. 2024, https://hapifhir.io. [Accessed: 01 Jyly 2024].
[42] Torab-Miandoab A, Samad-Soltani T, Jodati A, Rezaei-Hachesu P. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Med Informat Decis Making 2023;23(1):18.
[43] Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Dig Med 2019;2(1):79.
[44] Kieft R, Vreeke EM, De Groot E, Volkert P, Francke AL, Delnoij DM. The development of a nursing subset of patient problems to support interoperability. BMC Med Inf Decis Making 2017;17:1–12.
[45] Firely. Forge. 2024, https://fire.ly/products/forge/. [Accessed: 01 Jyly 2024].
[46] Khan W, Kumar T, Zhang C, Raj K, Roy AM, Luo B. SQL and NoSQL database software architecture performance analysis and assessments—A systematic literature review. Big Data Cogn Comput 2023;7(2):97.
[47] Blobel B, et al. CTS2 OWL: Mapping OWL ontologies to CTS2 terminology resources. In: PHealth 2022: proceedings of the 19th international conference on wearable micro and nano technologies for personalized health, vol. 299. IOS Press; 2022, p. 44.

Appendix 2

II

I. Bossenko, R. Randmaa, G. Piho, and P. Ross, “Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components,” *Frontiers in Digital Health*, p. 30, 2024

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 19 December 2024
DOI 10.3389/fdgth.2024.1480600

OPEN ACCESS

EDITED BY
Xia Jing,
Clemson University, United States

REVIEWED BY
Stefan Buttigieg,
University of Malta, Malta
Bernd G. M. E. Blobel,
University of Regensburg, Germany

*CORRESPONDENCE
Igor Bossenko
igor.bossenko@taltech.ee

RECEIVED 21 August 2024
ACCEPTED 13 November 2024
PUBLISHED 19 December 2024

CITATION
Bossenko I, Randmaa R, Piho G and Ross P (2024) Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components. Front. Digit. Health 6:1480600. doi: 10.3389/fdgth.2024.1480600

Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components

Igor Bossenko[1]*, Rainer Randmaa[1], Gunnar Piho[1] and Peeter Ross[2,3]

[1]Department of Software Science, Tallinn University of Technology (TalTech), Tallinn, Estonia, [2]Department of Health Technologies, TalTech, Tallinn, Estonia, [3]Research Department, East Tallinn Central Hospital, Tallinn, Estonia

Introduction: Ecosystem-centered healthcare innovations, such as digital health platforms, patient-centric records, and mobile health applications, depend on the semantic interoperability of health data. This ensures efficient, patient-focused healthcare delivery in a mobile world where citizens frequently travel for work and leisure. Beyond healthcare delivery, semantic interoperability is crucial for secondary health data use. This paper introduces a tool and techniques for achieving health data semantic interoperability, using reusable visual transformation components to create and validate transformation rules and maps, making them usable for domain experts with minimal technical skills.
Methods: The tool and techniques for health data semantic interoperability have been developed and validated using Design Science, a common methodology for developing software artifacts, including tools and techniques.
Results: Our tool and techniques are designed to facilitate the interoperability of Electronic Health Records (EHRs) by enabling the seamless unification of various health data formats in real time, without the need for extensive physical data migrations. These tools simplify complex health data transformations, allowing domain experts to specify and validate intricate data transformation rules and maps. The need for such a solution arises from the ongoing transition of the Estonian National Health Information System (ENHIS) from Clinical Document Architecture (CDA) to Fast Healthcare Interoperability Resources (FHIR), but it is general enough to be used for other data transformation needs, including the European Health Data Space (EHDS) ecosystem.
Conclusion: The proposed tool and techniques simplify health data transformation by allowing domain experts to specify and validate the necessary data transformation rules and maps. Evaluation by ENHIS domain experts demonstrated the usability, effectiveness, and business value of the tool and techniques.

KEYWORDS

FHIR Mapping Language (FML), TermX, semantic interoperability, data transformation, HL7 Clinical Document Architecture (CDA), HL7 Fast Healthcare Interoperability Resources (FHIR)

1 Introduction

Electronic Health Records (EHRs) are shared patient records that contain historical data about a patient compiled from all local Electronic Medical Records (EMR). EHRs serve a dual purpose in the healthcare ecosystem. Primarily, healthcare professionals use EHRs in healthcare delivery to access patient medical histories, diagnoses, treatments,

and treatment outcomes (1). Additionally, routine clinical data is valuable for secondary use in clinical research, public health assurance, healthcare financing, and health policy-making (2) by enabling the aggregation and analysis of health data to improve healthcare (3, 4).

The European Health Data Space (EHDS) initiative (5) aims to build a health data sharing ecosystem (6) within the European Union (EU), establishing standards, practices, infrastructures and governance to support the primary and secondary use of EHRs (7). It facilitates healthcare access across borders in a mobile world where people travel for work and leisure (8). While the EHDS has ambitious targets to improve data sharing and patient access across the EU, there are concerns that it might be too large an undertaking to succeed (9). Additionally, it could undermine patients' control over their data (10), complicate the work of healthcare professionals (9), and reduce public confidence (11). Furthermore, the challenges include inadequate compliance with existing regulations, such as the GDPR (12), potential excessive dominance and control by large tech companies (13), and deepening digital divides (14).

One possibility for adjusting the EHDS to more manageable goals with incremental steps is to utilize federated EHRs at different levels. These levels include the national level, such as the Estonian National Health Information System (ENHIS) (15), the healthcare institution level, such as in Austria where data is stored by the healthcare provider who first collected or generated it (16), and the citizen level, stored on citizens' devices (17). A more radical federation approach involves decentralized content-addressable storage networks fully owned and controlled by citizens (18). Federated EHRs, particularly at the citizen level, offer several benefits compared to those stored in unified data silos (17, 18):

- *Privacy and security*: Reduces the risk of large-scale data breaches by allowing patient data to remain within national borders.
- *Single points of failure*: Reduces the risk of single points of failure, enhancing system resilience.
- *Patient trust*: Ensures transparency and control over data sharing, encouraging greater patient engagement in healthcare initiatives.
- *Compliance with regulations*: Supports compliance with national and EU regulations, particularly the GDPR, by keeping data within jurisdictions and providing patients with control over their health information.

Despite strong security and data protection properties, federated EHRs face a major challenge: semantic interoperability (19), which involves creating a common understanding of data elements and their relationships, aligning data structures, and standardizing terminology. Different healthcare providers often use different standards and vocabularies, leading to inconsistencies and data integration and interpretation difficulties. Even with the same standards and vocabulary, differences in interpretation arise (20, 21), whether among software developers or domain experts, including physicians.

1.1 Research problem

The article addresses the need for the semantic interoperability of health data in various formats. The ENHIS, operational since 2008 and maintaining lifelong health records of all Estonian citizens (15), is transitioning from the HL7 Clinical Document Architecture (CDA) format to Fast Healthcare Interoperability Resources (FHIR) (22). To mitigate the risks associated with data migration, the system must operate with legacy CDA data while storing new data in FHIR format, necessitating on-the-fly semantic interoperability between both formats.

In addressing the specific real-world issue of converting CDA to FHIR, we framed it as a broader problem of transforming EHR data from one format to another in a semantically interoperable manner.

1.2 Research questions

This paper focuses on using reusable components to transform health data from CDA to FHIR, an approach which serves as a methodical basis for developing and modernizing health information systems toward seamless semantic interoperability. It contributes to achieving federated semantic interoperability rather than integrated (common data format) or unified (common standard) interoperability (23). Federated interoperability allows different systems to work together coherently and efficiently, enabling dynamic networking with minimal costs (24). Each system can use its preferred data transmission protocol internally, with adapters performing the necessary conversions based on specified transformation rules and maps. Our paper provides tools and techniques for creating these transformation rules and maps, enabling semantic data transformations on the fly.

A Dutch study (25) compared CDA and FHIR representations for the inter-convertibility and consistency of Detailed Clinical Models (DCMs). While most aspects were adequately represented, issues with restrictions, coded values, narrative structures, and attribute meanings could lead to semantic challenges, emphasizing the need for the right DCM implementation standards. Austrian (26), Italian (27), and Estonian (28) studies demonstrate the potential for transforming International Patient Summaries (IPSs) (29) from HL7 CDA documents to FHIR resources. However, these transformations were hard-coded (30), making them opaque to business analysts, difficult to reuse, rigid, and challenging to maintain long-term (31).

Our goal is to provide a robust and reliable health data transformation process that can be replicated and reused in various contexts, with two important objectives:

- *The problem of clarity*: Implementing a low-code/no-code pattern should facilitate the faster delivery of transformations by minimizing hand-coding and utilizing a graphical user interface. Visual representation should conceal the complexity of the data transformation language, enabling analysts to

adapt quickly. This strategy should increase efficiency and productivity and reduce dependency on developers.

- *The problem of reuse*: Reusing transformation rules and maps should save time and costs and improve efficiency, consistency, and readability. It should also lessen challenges such as initial investment, compatibility, and flexibility. Ensuring reusability requires careful planning and standardization. Visual representations can simplify understanding and apply complex transformations, while clear guidelines should facilitate reuse. This approach should enhance data processing quality and reduce the learning curve, fostering a more collaborative and efficient work environment.

Research rigor is centered on systematically developing visual mappings to facilitate data transformation. It emphasizes enhancing the clarity of transformations and promoting their reuse. This is demonstrated by customizing CDA and FHIR models, developing effective transformation rules and maps, and instantiating FML transformations.

1.3 Research results

Our work consolidates the experience of mapping and transforming data between HL7 CDA and HL7 FHIR R5 within the Estonian National Health Information System.

Using a Design Science (DS) methodology (32), we developed techniques for domain experts to create and reuse visual health data transformation components, along with preliminary techniques for ensuring their correctness.

After analyzing existing data transformation languages and tools, we support the use of the FHIR Mapping Language (FML). To address the lack of suitable tools for domain experts (33), we designed, developed, and validated the TermX tool (34, 35) with input from domain experts (36, 37). TermX allows domain experts to specify and test transformation rules and maps between data formats using a WYSIWYG[1] approach with minimal technical knowledge (38).

1.4 Outline of the paper

The paper is organized as follows: Section 2. explains the HL7 CDA to FHIR transformation challenges, the TermX tool we developed for data transformations, and the methods we use in creating the data transformation techniques. Section 3. documents the transformation techniques. Section 4. evaluates the proposed techniques and discusses the related social impacts in the context of the EHDS. It also discusses related work, including an analysis of the pertinent tools and languages. Finally, in Section 5, we conclude and outline directions for future research.

[1]What you see is what you get.

2 Methods

We aim to improve data transformations by designing techniques and reusable WYSIWYG transformation components that domain experts can use to specify and validate data transformation rules and maps for semantic interoperability in EHR infrastructure, with only minimal technical expertise and skill needed. We adhere to the Design Science (DS) methodology (32, 39). A *transformation rule* is a specific instruction or set of instructions that defines how a particular piece of data should be transformed (40). A *transformation* refers to the overall process of converting data from one format or structure to another (40). A *transformation map* is a set of transformation rules and metadata used by the transformation engine during the transformation process (41). A *transformation component* is a visual representation of a transformation rule or map in TermX Visual Editor that contains an FML code that makes the necessary transformations. The techniques and transformation components, along with the TermX tool we use, are our artifacts. The context of these artifacts in performing health data transformations is the IT infrastructure of health organizations and state agencies. DS problems are improvement problems. This work aims to improve the federated semantic interoperability between heterogeneous healthcare EHRs. The proposed techniques are illustrated with data transformations from CDA to FHIR.

DS is part of the engineering cycle (Figure 1) and includes the problem investigation, treatment design, and treatment validation phases. The treatment implementation phase is not part of DS but forms an engineering cycle along with the DS phases. This paper reports two DS cycles and therefore also two engineering cycles. In the first cycle, we designed and developed the TermX tool. In the second cycle, we evaluated the TermX tool by designing the techniques and reusable WYSIWYG components for data transformation rules and maps from CDA to FHIR.

While the implementation of the artifact (TermX tool) is not part of DS but part of the engineering cycle, Figure 1 includes its implementation to illustrate the place and role of the TermX tool's development in our study. We designed TermX according to the DS methodology, encompassing the following steps: (1) investigating a problem, problem relevance, and research rigor by reviewing published papers on existing data transformation languages, tools, and implemented projects (see Section 4.1); (2) designing the TermX tool (38); and (3) validating the TermX design with domain experts from various countries (see Section 2.2).

In the second cycle, the main focus of the current paper is to evaluate the TermX tool by designing visual reusable transformation components that domain experts can use for CDA to FHIR transformations. We also generalize the transformation components' development process as techniques for developing reusable transformation components using TermX (Section 3) and explain the relevance of our research in the EHDS ecosystem, including how the proposed approach supports federated semantic interoperability (Section 4).

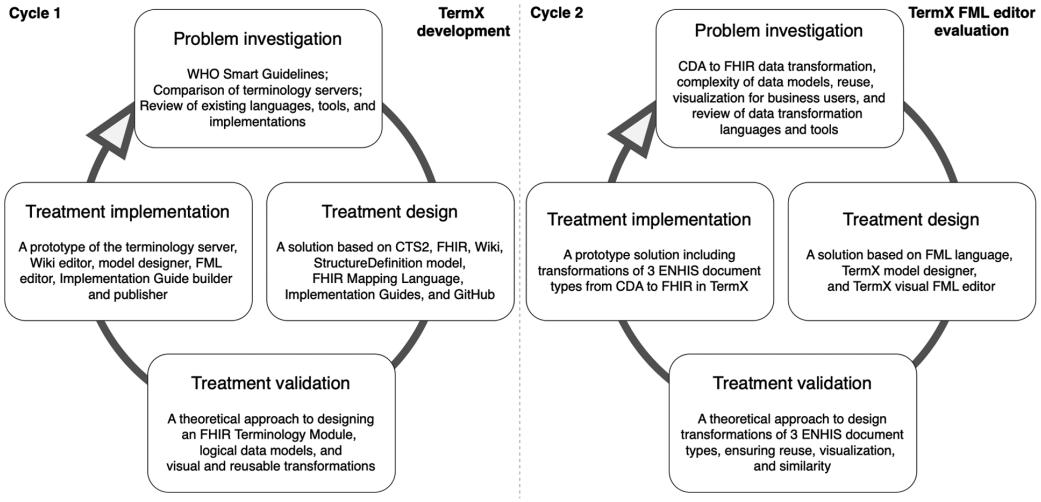

FIGURE 1
The Design Science methodology used for the development and evaluation of TermX.

2.1 HL7 CDA to FHIR transformation

HL7 CDA (42) and HL7 FHIR (43) are two widespread standards for the interoperability of health information systems. Although these two standards are designed to be interoperable, the semantic heterogeneity of various software vendors' implementations inhibits semantically correct model transformations between these standards (44). Additionally, model transformations between specific HL7 CDA and HL7 FHIR implementations are not straightforward and there is no single correct way to achieve them (27). Therefore, highlighting a new tool and the related techniques is pertinent, as transformation techniques between CDA and FHIR are relatively undocumented in academic literature.

HL7 CDA is a template-based and XML-centric standard for health data documents, first released in the early 2000s (42). It is a complex standard with many shortcomings in data redundancy and analysis. HL7 FHIR, by contrast, is a modern interoperability framework based on widespread web technologies, such as REST and JSON (44, 45). The shortcomings of HL7 CDA have been largely addressed in FHIR, which is why mapping and transforming existing HL7 CDA formatted health data to HL7 FHIR resources in a semantically interoperable way has tremendous potential and value in both health data usage and health data analysis-related innovation (46).

Although CDA and FHIR are designed to be interoperable, both standards are complex, and transformation between them is non-trivial (46). For example, the HL7 Reference Implementation Model (RIM) used within HL7 V3 and CDA aims to encompass the full spectrum of possible healthcare scenarios (47). In contrast, HL7 FHIR provides a model for the most common scenarios. Instead of defining a complete model for all aspects of healthcare, FHIR follows the 80/20 principle by defining only the most common health scenarios, adding the possibility of extension to cases where customization is necessary (48, 49).

The FHIR authors have identified various interoperability challenges when transforming data from CDA format to FHIR. Key points include clinical content mapping at the template level, managing differences in narrative granularity, and handling discrete-to-human-readable linkages, with some potential information loss when converting from CDA to FHIR (50). Additionally, both CDA and FHIR standards have evolved over time, and each new version brings changes that may not be compatible with previous versions (51–53). Efforts also exist to maintain forward and backward compatibility between versions, which is not guaranteed in all cases (53).

It is important to note that while CDA and FHIR are specifications for health data exchange, they differ in their approach and usage. FHIR's resource-based model allows for more granular control and flexibility, whereas CDA's document-centric approach provides a robust and standardized format for clinical documents. They also differ in their licensing requirements: CDA requires a license for use, whereas FHIR is dedicated to the public domain to encourage widespread adoption.

2.2 TermX: a game changer in interoperability

The necessity of robust, enduring, and relevant healthcare interoperability is universal across all clinical and health domains. However, we identified a gap in the availability of open-source, cost-free, high-quality tools that offer multilingual support and an advanced graphical interface (33). To address

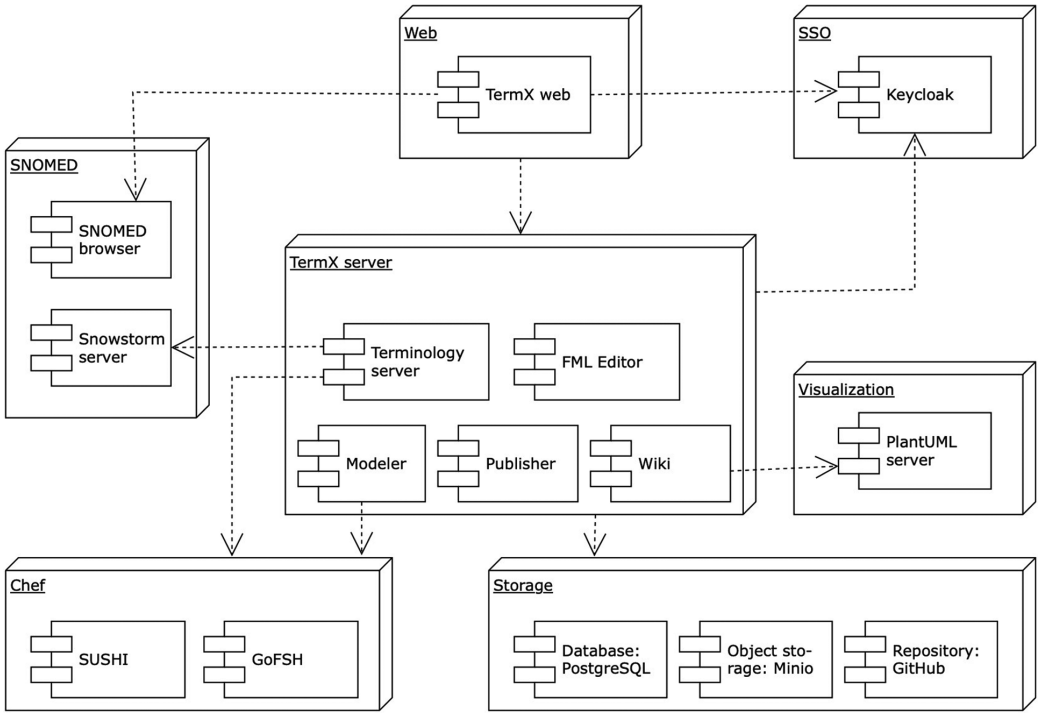

FIGURE 2
TermX component diagram.

this, we designed and implemented TermX – a novel, open-source platform for terminology management and data transformations to support interoperability between healthcare institutions and systems (34). TermX incorporates a terminology server, a Wiki, a model designer, an FML transformation editor, and tools for authoring and publishing (35). Figure 2 visualizes the TermX components (38). TermX is designed to manage data models and transformations and develop terminology and implementation guides for healthcare systems at international, national, regional, and hospital levels. It aims to ensure open, standardized access to published data and guarantee semantic interoperability based on the FHIR standard. We have validated TermX with TalTech (Tallinn University of Technology, Estonia), the private sector, and national standardization agencies in Estonia, Lithuania, Uzbekistan, and the Czech Republic.

TermX provides a visual model designer and FML Editor for creating and visualizing data models and FML transformation rules and maps through a user-friendly interface (Figure 3). They are designed specifically for business analysts rather than developers. The model designer implements the FHIR StructureDefinition specification (54) and provides the capability to manage data models through a user-friendly interface or formal specification in FML code. The FML editor's core purpose is to design transformation components, hide the complexity of the CDA, FHIR, and FML languages, and enable analysts to adapt quickly.

TermX uses the FHIREST (55) and HAPI FHIR (56) libraries to provide the FHIR API and uses HAPI FHIR (57) as the foundation for its transformation engine, transforming data from input sources into output sources (38). TermX was created as the result of an academic project at TalTech.

2.2.1 Reusable visual transformation components

CDA and FHIR are health data interoperability models developed by HL7 (44); both are designed with a hierarchical structure of data types and resources. For instance, CDA includes four code data types: CS (code simple), CV (coded value), CE (code with equivalents), and CD (concept descriptor) (see Figure 4). CS is the simplest, while CD is the most complex. Complex data types are composed of simple data types. In CDA, the simplest data type may be a subset of a more complex data type, for example, a CS is a subset of a CV data type. In FHIR, resources are categorized into metadata, special-purpose, general-purpose, and primitive data types (58). In both models, the depth of objects in the XML or JSON document tree can become very large. In the case of large CDA documents, the depth of the document trees results in very voluminous transformations.

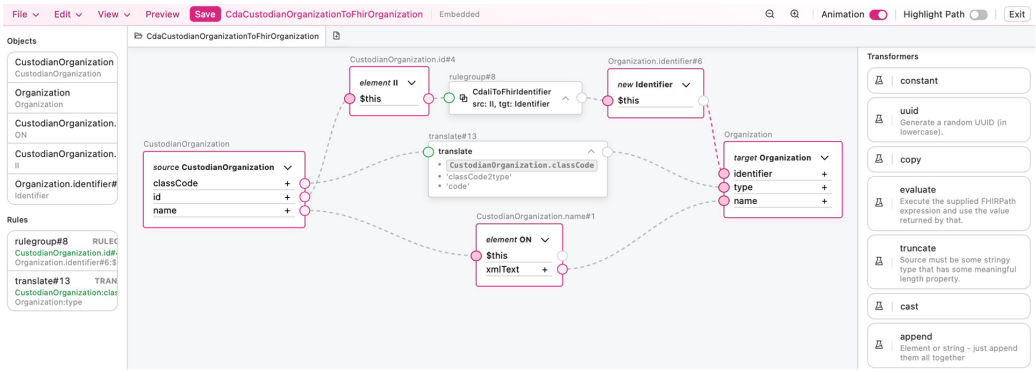

FIGURE 3
User interface of the TermX FML editor.

The transformation is the entire process of converting the resource, while the transformation rules are the detailed steps that specify how each attribute within the resource should be handled. Transformation rules are applied to convert the extracted data from its previous form into the required form. These rules could involve various instructions, such as extraction, conversion, or formatting. The transformation map, conversely, is not just an abstract concept but manifests itself as a tangible artifact. Every transformation map may be reused as a transformation rule in another transformation. Correct transformation rules and maps are fundamental in defining transformations, such as transforming CDA documents to the FHIR Bundle resource (59), as needed in the ENHIS. We identified the required transformation rules and maps between the data types and models of these two standards and created corresponding transformation components. We found that transformation components of simple data types, such as CD to CodeableConcept and II (instance identifier) to Identifier (see Figure 4), can be reused in more complex data types and model transformations. Such reuse simplifies the development of transformation rules and maps, improves clarity, and reduces the needed FML source code.

2.3 Research towards reusable visual transformation techniques

2.3.1 Problem investigation

The data transformation from CDA to FHIR necessitates a profound comprehension of the data structures inherent in both standards. FHIR *StructureDefinition* (54) describes a resource structure and defines a set of data element definitions and associated usage rules. These structure definitions describe the content defined in the FHIR specification, such as resources, data types, and underlying infrastructural types, and how these structures are utilized in implementations.

In CDA, each element is comprehensively defined using standard schema definition (XSD) files. These XSD files act as architectural designs, delineating the structure and data types of CDA documents and simplifying the process of validating these documents against the prescribed schema. The CDA model is based on the HL7 Reference Information Model (RIM) and utilizes reusable data types, templates, sections, and components (50). For instance, patient demographics, medication information, and clinical observations are standardized and reused across different CDA documents. HL7 has implemented a representation of the CDA R2.0 specification using FHIR Logical Models expressed as FHIR StructureDefinition instances available under an open-source license (60).

Many models in CDA and FHIR have numerous attributes, are complex, and contain hierarchies. We need a way to reuse data type transformations and provide reusable transformation components for CDA and FHIR subtypes, such as CD to Coding and II to Identifier. This approach will enhance the efficiency and reliability of data-handling processes. For instance, the ENHIS "Outpatient Case Summary" comprises 24 sections, while the "Birth Summary" comprises 17 sections (61). Of the "Birth Summary" sections, only four are absent in the "Outpatient Case Summary". Our techniques involve creating transformation components for a single document type and then applying these components to different types of documents. If new sections are introduced in the new document type, transformation components are only developed for these new sections and included in the reusable transformation components library. With each new document type, the number of sections requiring transformation components development will decrease and eventually reach zero. We also need a solution to validate transformation components to identify problems during development rather than production and to avoid errors during the development of transformation components.

Transformations of simple data objects are straightforward, and the associated source code in FHIR Mapping Language is relatively uncomplicated. However, with the transformation of hierarchical complex objects, the source code becomes highly intricate and may pose comprehension challenges for domain experts. Complex transformations necessitate visualization (62).

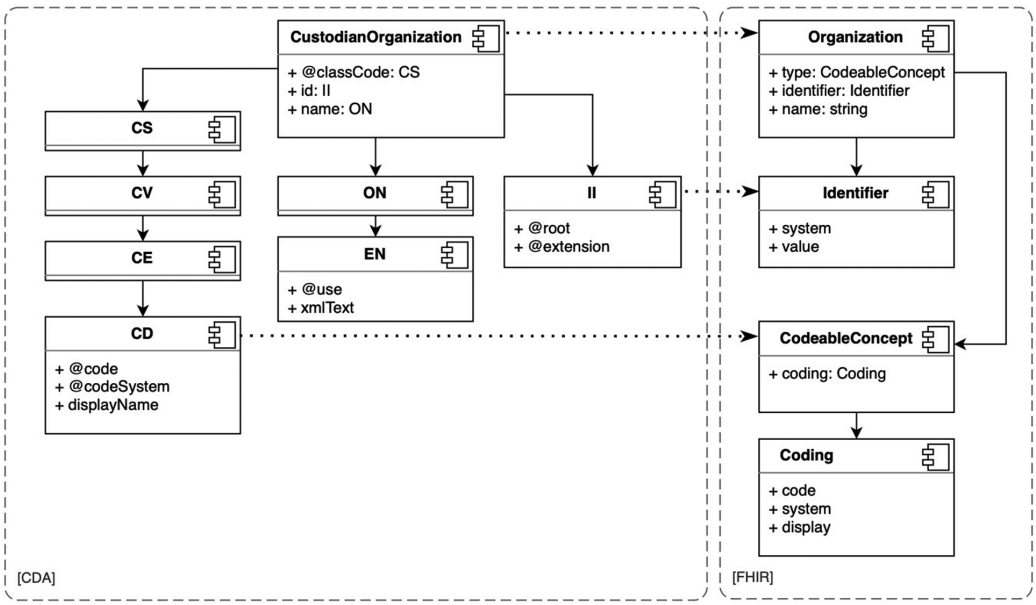

FIGURE 4
CDA CustodianOrganization and FHIR Organization resources, subtypes and related mappings.

We aim to establish a set of CDA and FHIR transformation components encompassing a broad spectrum, ranging from primitive data types to complex resources, and formulate appropriate techniques. We hypothesize the following:

(1) TermX as an artifact will apply to all CDA data types, sections, and documents.
(2) All transformation components can be developed using the TermX visual user interface.
(3) The developed transformation components can be reused.

This strategy would facilitate the reuse of prior transformation components, thereby augmenting the efficiency and uniformity of transformation procedures. Such an approach is designed to fortify the robustness and adaptability of the developed TermX tool, equipping it with the capacity to help domain experts develop and validate transformation components by hiding the details and complexities embedded within CDA and FHIR data models.

2.3.2 Treatment design

Based on the problem investigation above, we have established the following requirements for the visual reusable transformation components set:

(1) It must support strict data models
(2) It must support the reuse of transformation components
(3) It must have native support for CDA and FHIR
(4) It must support the WYSIWYG approach

This approach underscores our commitment to advancing the field of data transformation and management, ensuring that our data transformation techniques are accessible and understandable to a broad range of stakeholders.

The selected approach evaluates the usability of the TermX model designer and the TermX visual FML editor, the FML language, and the HAPI FHIR implementation of FML used by TermX (Figure 2 illustrates the TermX architecture and components). TermX enables the registration of HL7 V3 and CDA models in the TermX model designer, uses FHIR resource definitions, creates data transformation rules from CDA to FHIR in the TermX visual FML editor, and publishes the transformations on GitHub.

The transformation may be triggered by HTTP requests within scripts or through the web user interface. TermX is available as a set of Docker containers used for deployment. We use the logical models provided with the HL7 CDA R2.0 core standard (60) as a basis for ENHIS CDA input instances. These models were extended according to the ENHIS CDA standard implementation. We used FHIR Release 5 (R5) structure definitions (54) as the standard for output instances. The transformations handle one input CDA file and output one FHIR file.

2.3.3 Treatment validation

Treatment validation ensures that the chosen approach contributes to achieving stakeholders' goals when implemented. Our approach includes prototyping a set of transformation components using ENHIS version 8.2 CDA documents, the

FHIR R5 specification, and the TermX tool. The FML Editor achieved Technology Readiness Level (TRL) 5 according to the European Commission's classification (63) at the start of the validation process. The dataset, derived from three ENHIS HL7 CDA document types: the "Outpatient Case Summary", the "Notice of Growth", and the "Birth Summary", was established during the research to validate the proposed transformation techniques. For each selected document type, we used a sample CDA document from the ENHIS specification that includes all available sections.

The ENHIS "Outpatient Case Summary" includes 24 data sections, the "Notice of Growth" includes seven sections with two unique sections, and the "Birth Summary" includes 17 sections with four unique sections. Initially, we developed transformation components for all sections in the "Notice of Growth" and their associated classes and data types. Additionally, we created a transformation component to convert the "Notice of Growth" document into FHIR, incorporating all the transformations in the created section. For each subsequent document, we created a new transformation component that included the transformation components of the existing sections. Then, we added new section transformation components and linked them to the particular document transformation component. With the implemented prototype, we successfully verified that: (1) TermX was applicable for all necessary CDA data types, sections, and documents; (2) all transformation components were developed using the TermX visual user interface; and (3) the developed transformation components were reused in subsequent data types, sections, and documents.

The results obtained were first validated manually by comparing CDA and FHIR messages section by section to ensure the correctness of transformations. Next, we designed a technique (Section 2.3.4) to automate the validation process. Subsequently, the results were demonstrated to the IT department of the Health and Welfare Information Systems Centre (TEHIK), which operates the ENHIS. The feedback was overwhelmingly positive, with the team expressing their approval and satisfaction. Following the internal evaluation, TEHIK chose it as their transformation tool.

2.3.4 Advance techniques for validating transformation rules

Transformation validation should be deterministic, with each transformation having a dedicated test suite using predefined human-validated inputs and expected outputs. While developing these deterministic input-output pairs is time-consuming and can lengthen the development cycle, it is essential for robust production solutions and sometimes required by legislation (64, 65). We envision quicker heuristic feedback techniques for prototyping or experimentation, combining FHIR structure validation and an input-output content similarity assessment using a natural language processing (NLP) solution. However, supporting dedicated test suites in TermX and developing these heuristic validation techniques will largely be a part of future work.

Data similarity between the original HL7 CDA and the transformed HL7 FHIR documents was validated. No specialized out-of-the-box tool capable of statistically evaluating the correctness of the transformations was found. Therefore, CDA and FHIR documents were converted into collections of key-value pairs to which statistical tools were applied (66). The highest similarity percentage was achieved using the Term Frequency-Inverse Document Frequency (TF-IDF) methods (67). Further research in this direction is planned for the future.

3 Development techniques for reusable visual transformation components

Our study results in developing hierarchical, reusable transformation components for converting CDA documents into the collections of FHIR resources [Bundle (59)]. It highlights techniques that use the FHIR Mapping Language and the TermX visual editor to improve reuse and clarity in data transformations. First, we introduce the devised techniques. Then, we illustrate how the visual TermX editor supports our approach, making it accessible to analysts through a no-code visual interface. We provide practical examples using the ENHIS CDA documents, specifically the "Notice of Growth", "Outpatient Case Summary", and "Birth Summary", to demonstrate the application of these techniques in real-world scenarios. Furthermore, we outline preliminary techniques for validating transformation components, emphasizing the need for deterministic testing and proposing heuristic feedback techniques.

3.1 Techniques for hierarchical reusable transformation components

According to the authors of FHIR, transformations from CDA to FHIR should be performed at the template level (50). A CDA template follows a specific structure: the entire document is encapsulated within a <*ClinicalDocument*> element, which includes header information and a <*structuredBody*> element. The <*structuredBody*> element is composed of <*component*> elements, which in turn consist of <*section*> elements (Figure 5). These <*section*> elements comprise standard HL7 CDA classes, with optional extensions defined by the implementer. CDA classes are assembled using other CDA classes and complex and primitive data types. FHIR resource definitions also use other definitions and data types. A transformed CDA document is presented as an FHIR Bundle—a container holding a collection of FHIR resources.

We propose that the issues of reuse and clarity in CDA to FHIR transformations can be addressed through a hierarchy of reusable transformation components organized similarly to the structure of a CDA document. The FHIR Mapping Language allows the reuse of transformation maps that can be invoked from other transformation rules, thereby supporting our proposed approach.

We commence by delineating a hierarchical structure of data types and models. This hierarchy is instrumental in encapsulating the complexity and diversity of healthcare data. The fundamental

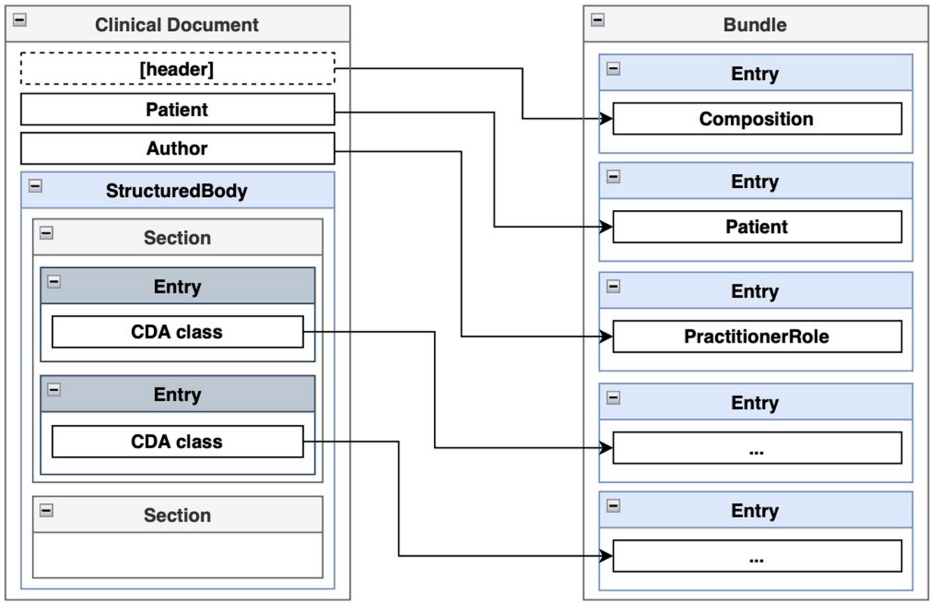

FIGURE 5
Mapping of ClinicalDocument to FHIR Bundle.

units can be categorized into primitive, basic, and complex data types. Each of these categories represents a different level of abstraction and complexity. Primitive data types are the simplest and most fundamental, representing basic data elements such as strings and numbers. Basic data types are slightly more complex, encapsulating the related data elements. Complex data types, on the other hand, represent a collection of basic and primitive data types, forming a more intricate structure. Subsequently, we establish transformation components between these data types. These transformation components elucidate the relationships and transformations between data types, thereby facilitating interoperability and data exchange. Lastly, we construct transformation components between different models.

Our findings demonstrate that it is feasible to define reusable transformation components at various levels of granularity of a CDA template: the complex data type level, the CDA class level, the section level, and the document level. The primitive data types between CDA and FHIR are already interoperable. Based on these levels of granularity, we establish sets of transformation rules to be maintained.

With the different granularity level transformation components, a set of *ConceptMap*, and the source and target *StructureDefinitions*, we define a set of software artifacts to be created and maintained for developing robust CDA to FHIR transformation components quickly. The list of artifacts is described in Table 1, and the dependencies among the artifacts are visualized in Figure 6. We designed the transformation components to transform basic and complex data types from CDA to FHIR. Mappings from CDA sections to FHIR resources are assembled using CDA class to FHIR resource transformation components and CDA complex data type to FHIR complex data type transformation components. Subsequently, the CDA document for FHIR bundle transformation components can be formed using the CDA section for FHIR resource transformation components. The CDA document header is considered a section in our approach. Lower levels of granularity transformation components are used in the transformation components with the higher granularity level, thus adhering to one-way dependencies—an important software architecture pattern.

In addition to these transformation components, two additional components are required. The *ConceptMap* (68) translates the set of concepts in one code system to one or more concepts in other code systems. The *StructureDefinitions* (54) are used to define source and target data models of the transformations.

The reuse problem is addressed using a single transformation component in multiple other transformation components where the same construct is mapped. For example, a component that maps a CDA *II* class to a FHIR *Identifier* data type can be used in components mapping both the CDA class *CustodianOrganization* to the FHIR *Organization* resource and the CDA class *AssignedAuthor* to the FHIR *Practitioner* resource. By solving the problem of reuse, we ensure that issues in transformations have a single point of failure, thereby enhancing the robustness of the transformations. Reuse also enables the faster development of transformation components from CDA templates to FHIR bundles, as it eliminates the need to repeatedly write the same transformation component for transforming the same section or class to FHIR when working with different CDA templates.

TABLE 1 CDA2FHIR artifacts.

Artifact	Source	Target	Explanation
I/O structures			The definitions of the structures for the inputs and outputs of the transformations in the form of FHIR StructureDefinition resources.
Classifier mappings			FHIR ConceptMap resources that map CDA coding systems to FHIR coding systems.
Data type Mappings	CDA data type	FHIR data type	Transformations between CDA data types and FHIR data types in the form of FML or FHIR StructureMap resources.
Class to Resources Mappings	CDA class	FHIR resources	Transformations between CDA classes and FHIR resources in the form of FML or FHIR StructureMap resources, constructed from the elements of data type transformations and classifier transformations.
Section to Resources Mappings	CDA *<section>*	FHIR resources	Transformations between CDA document sections and FHIR Bundle resources in the form of FML or FHIR StructureMap resources. A document section is a code-distinguished section within the structuredBody element of a CDA document or the CDA document header. These transformations are constructed from the elements of transformations between CDA classes and FHIR resources as well as data type transformations.
Document to Bundle Mappings	CDA document template	FHIR bundle	Transformations between CDA documents and FHIR Bundle resources in the form of FML or FHIR StructureMap resources. These transformations are constructed from the elements of transformations between CDA document sections and FHIR Bundle resources.

The problem of clarity is addressed through reusable transformation components that encapsulate complexity at various levels of granularity. When analyzing a component that transforms a CDA template to an FHIR bundle using our proposed techniques, we only need to understand the different sections defined in the template without being burdened by the details of the transformation component of CDA classes or complex data types. This principle applies to rules at each level of granularity, ensuring that each component remains focused and comprehensible by abstracting lower-level details.

3.2 Techniques for visualizing transformation components with TermX

To support the described techniques for developing CDA to FHIR transformation components using FML, a visual editor must support the following use cases: the management of *StructureDefinitions*, the management of *ConceptMaps*, the creation of FML transformation, and the ability to use existing transformation components in other FML transformations. According to our results, the TermX software supports all of these use cases through a visual user interface with low-code/no-code.

In TermX, the management of *StructureDefinitions* is part of the Modeler module. *StructureDefinitions* can be displayed as a tree-like visual structure and edited without modifying the underlying JSON or FHIR Shorthand (FSH) (69) source. Additionally, the HL7 CDA *StructureDefinitions* do not need to be implemented from scratch, as the FHIR authors have provided multiple core standard CDA specifications using FHIR Logical Models expressed as FHIR *StructureDefinition* instances (60). These logical models can serve as a basis for *StructureDefinitions* of a specific CDA implementation. The CDA *StructureDefinitions* can be created in TermX using the provided JSON or FSH syntax and then edited with the visual editor to fit specific implementation guidelines. A FHIR implementation generally includes an *Implementation Guide* containing the Resources' *StructureDefinitions*.

The Terminology module supports the management of *ConceptMaps* that represent the mapping between source and target terminology. The *ConceptMaps* can be used as a *transformation rule*.

TermX provides a visual FML editor as a designer of explicitly designed FML transformation components for business analysts (38). Every transformation has at least one source and target *StructureDefinition* and may reuse other FML transformation components and *ConceptMaps*. The imported elements can then be utilized on a visual canvas, dragging and dropping as boxes. Lines can be drawn between the boxes, visually modeling the control flow of the transformation rule from the source structure to the target structure, from which FML code is generated (Figure 3). The objective of the FML editor is to visually represent transformation rules, hide the complexity of the FML language, and facilitate rapid adaptation to the FML language.

In the work described in this paper, all the necessary transformation components were created with the visual editor of TermX; even the code generated behind certain transformation component visualization boxes and lines was not always intuitive to inexperienced users.

3.3 Techniques for developing CDA to FHIR transformation components

We evaluated the viability of the proposed techniques by developing a prototype development for transforming the ENHIS CDA documents "Notice of Growth", "Outpatient Case Summary", and "Birth Summary". We began by dividing the "Notice of Growth" into sections and then breaking those sections into classes and data types. We also documented the necessary *ConceptMaps* and *StructureDefinitions*. After this, we developed the transformation components, starting with lower granularity artifacts. This process was repeated for the other two CDA documents, reusing already specified transformation components wherever possible. Subsequently, we provide

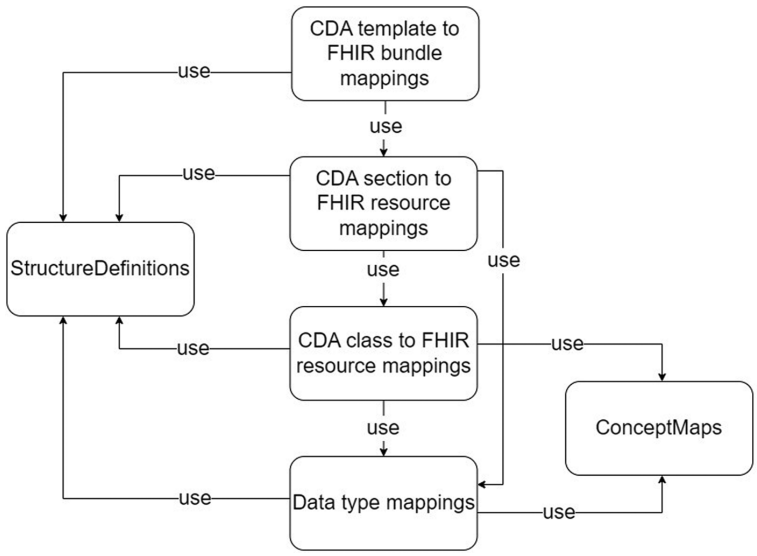

FIGURE 6
Usage dependencies between artifacts used in CDA to FHIR mappings.

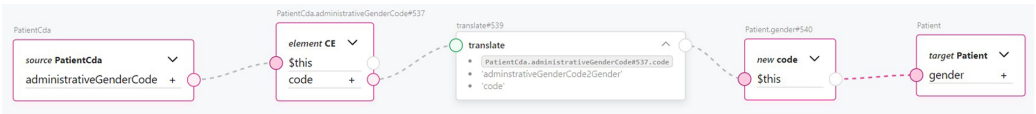

FIGURE 7
Transformation of CDA administrative gender attribute to FHIR gender attribute using ConceptMap for concept translation.

examples from a real-world use case to illustrate the key points previously highlighted.

3.3.1 Specifying CDA data type level transformations

For the ENHIS CDA *StructureDefinitions*, we were able to use the logical models provided with the HL7 CDA R2.0 core standard (60) as a basis, which were then modified as needed according to the ENHIS CDA standard implementation. This implementation is available as Enterprise Architect models and PDF documents on the web and is accessible within the Estonian IP address space. The modifications required for the core standard *StructureDefinitions* were necessary to address the extensions of the base model defined in the Estonian implementation as well as instances of misuse of the standard. For example, in the CDA *Observation* class, the *Ratio* data type for the value attribute is denoted as *RTO-PQ-PQ* in the core standard, which employs hyphens. However, in the ENHIS implementation, it is referred to as *RTO_PQ_PQ*, where underscores are used instead. An example of an extension that needed to be accounted for is the *<asLicencedEntity>* element added to the *<assignedEntity>* element to provide information about the authority licensing the healthcare worker. As the transformation target structure, we used the base FHIR R5 release, for which we utilized URIs in a test server.

An example of using *ConceptMaps* and terminology translation between CDA and FHIR is illustrated when transforming the CDA Patient class into the FHIR Patient resource. The two standards use different sets of codes to represent the administrative gender of the patient. For instance, in the ENHIS CDA implementation, the code "N" represents the female gender, whereas in FHIR R5, the code "female" is expected. A *ConceptMap* was constructed and used with the transformation rule to perform translation between the two terminology code systems, as shown in Figure 7. In the figure, the *administrativeGenderCode* attribute of the Patient CDA class is piped into the transformation rule, the result of which is assigned to a new FHIR code data type and then to the gender attribute of the Patient FHIR resource.

One of the most common transformations we encountered was between the FHIR concept and different representations of the CDA

concepts. For example, FML transformation rules between the CDA CD class and the FHIR *CodeableConcept* resource as well as between the CDA CE class and the FHIR *CodeableConcept* resource provided significant value in terms of reuse. These transformation rules were very common in higher granularity level transformations. Due to the nested structure of the FHIR *CodeableConcept* and the three data attributes mapped between the structures, calling a reusable transformation rule with one line of code saved us from repeating the same six lines of code each time. An example of a reusable CDA CE to FHIR *CodeableConcept* transformation rule using the TermX visual editor can be seen in part A of Figure 8. The attributes of the CE CDA class are assigned to a new Coding FHIR resource. The Coding resource is then assigned to the target *CodeableConcept* coding attribute. Specifically, the CE CDA class's code attribute corresponds to the FHIR Coding's code attribute, the *codeSystem* attribute corresponds to the *system* attribute, and the *displayName* attribute corresponds to the *display* attribute.

Notably, FML also enabled us to handle semantically faulty XML at the data type level. In an *Observation* element in the "Outpatient Summary" test documents we used, we encountered a decimal value represented as text with a comma decimal separator inside an *EncapsulatedData* data type: *<value xsi:type="ED">12,2</value>*. To fix this issue, we were able to replace the decimal separator and cast the text into a decimal data type using FML's *evaluate* rule with a *FHIRPath* expression and a *cast* rule. We accomplished all of this using only the visual editor (see Figure 8 part B). The inner text of the XML tag represented by the *xmlText* attribute is piped into an *evaluate* block, where a *FHIRPath* expression is used to replace the comma with a period in the text string. The evaluated string is piped into a *cast* block, which casts it to a decimal data type and assigns it to an output value. In our opinion, this result illustrates that a visual editor can produce fault-tolerant and robust transformation rules.

3.3.2 Specifying CDA class level transformations

CDA class to FHIR resource transformation rules can be exemplified with Figure 9, which shows how a CDA *AssignedAuthor* class is mapped to a FHIR *Practitioner* resource using the TermX visual editor. The CDA *AssignedAuthor* class is split into the *II* data type from the id attribute, the CE data type from the code attribute, and the Person class from the

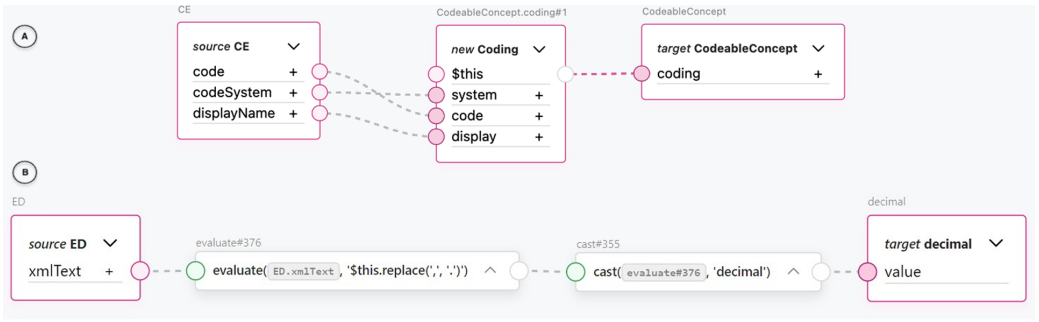

FIGURE 8
An example transformation from CDA CE class to FHIR CodeableConcept resource.

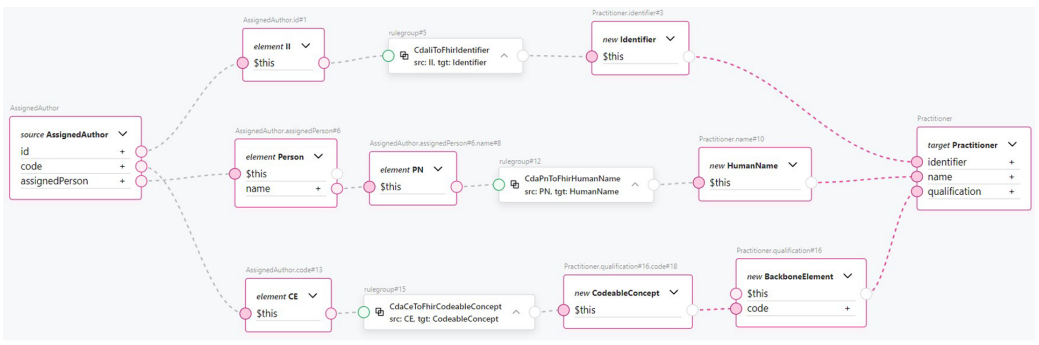

FIGURE 9
An example transformation from CDA AssignedAuthor class to FHIR Practitioner resource.

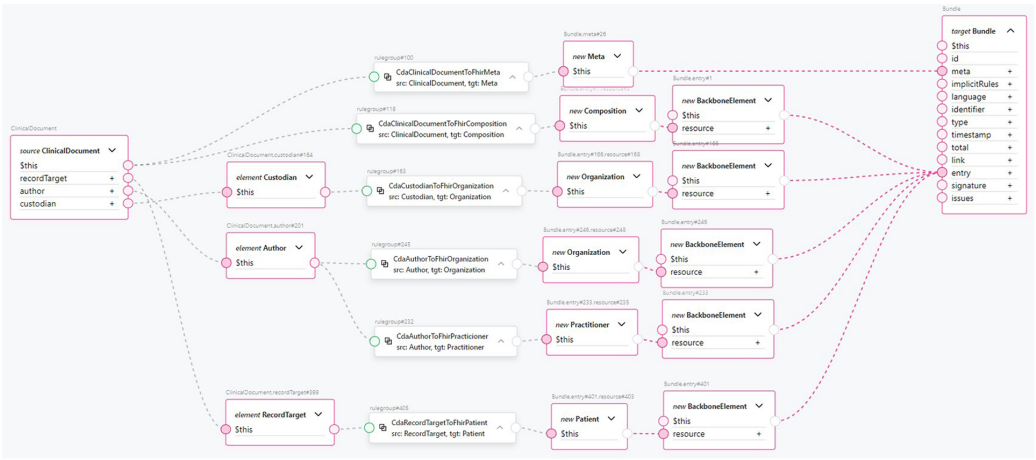

FIGURE 10
An example transformation from CDA ClinicalDocument header entries to FHIR Bundle entries.

assignedPerson attribute. Subsequently, the CDA *II* data type is transformed into the FHIR *Identifier* resource using the reusable transformation component *CdaIiToFhirIdentifier*. The CDA *CE* data type is transformed into the FHIR *CodeableConcept* resource using the reusable transformation component *CdaCeToFhirCodeableConcept*. The CDA *PN* data type is extracted from Person class and transformed into the FHIR *HumanName* data type using the reusable transformation component *CdaPnToFhirHumanName*. The transformed FHIR resources are then assigned to the target *Practitioner* resource's identifier, qualification, and name attributes, accordingly. Notice how data type transformation rules are imported and then used. Referring to Figure 8, which shows the implementation of the *CdaCeToFhirCodeableConcept* transformation, it is clear how our approach encapsulates complexity and promotes clarity at the CDA class to FHIR resource mapping level.

3.3.3 Specifying CDA section level transformations

Transforming the CDA document header to FHIR is an example of the transformation component from a CDA section to a FHIR resource. This is shown in Figure 10. The clinical document header contains a variety of information. The confidentiality codes, as top-level attributes of the header, are transformed into FHIR's Meta resource and assigned to the FHIR Bundle's meta attribute. The structural information about the sections in the document is compiled to form the FHIR *Composition* resource and added to the FHIR Bundle as an entry. The clinical document header's *custodian* attribute, a CDA *Custodian* class instance, is transformed into a FHIR *Organization* resource and added to the bundle as an entry. The *author* attribute of the clinical document, a CDA *Author* class instance, contains information about the author's person and organization. Therefore, two transformation components are used: one for transforming the data into a FHIR *Organization* resource and another for transforming the data into a FHIR *Practitioner* resource. Both resources are added to the FHIR Bundle as entries. Finally, the *recordTarget* attribute of the clinical document header, a *RecordTarget* CDA class instance, is transformed into a FHIR *Patient* resource and added to the FHIR Bundle as an entry. This concludes the scope of our *ClinicalDocument* header transformation component. The number of transformation components is approximately equal to the number of document types and CDA classes used in them, considering the CDA class hierarchy. By encapsulating transformation components such as *CdaCustodianToFhirOrganization*, *CdaAuthorToFhirOrganization*, *CdaAuthorToFhirPractitioner*, and others into reusable transformation components, the CDA header transformation rule remains comprehensible, even though the amount of information to be transformed is much larger.

3.3.4 Specifying CDA document level transformations

Finally, using CDA section transformation components, we compose a transformation component for the "Notice of Growth" CDA document (see Figure 11). We find a document section by section code, then apply a reusable component to transform this section into FHIR resources, and then combine them into a FHIR Bundle. The header section is extracted from the root level of the *ClinicalDocument*, while the other sections are extracted from within the <*StructuredBody*> element. From the <*structuredBody*> element, we extract two sections: the *AGE* section and the *GROWTH* section. The *AGE* section is transformed into an *Observation* FHIR resource containing the patient's age information using a single *CdaAgeSectionToObservation* reusable transformation component.

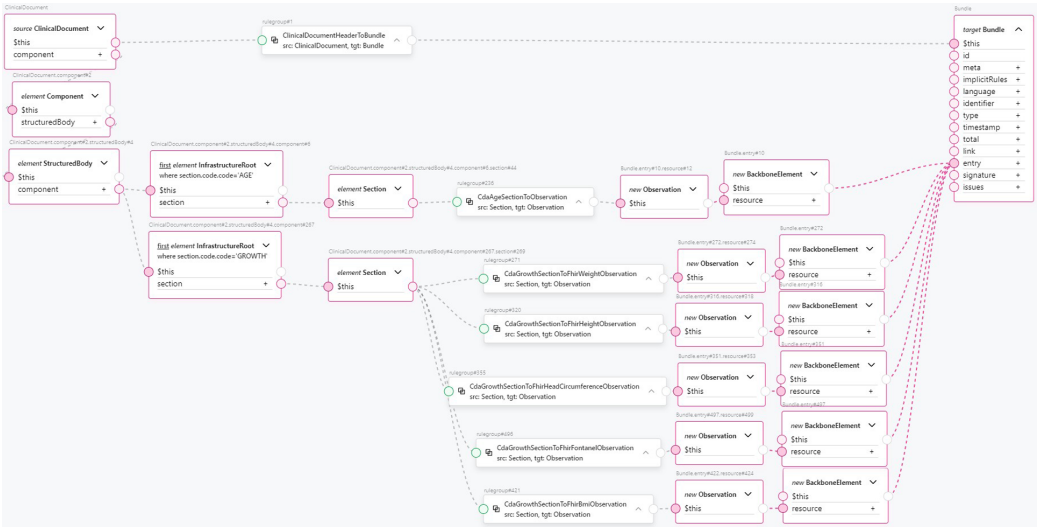

FIGURE 11
An example transformation from CDA Growth Report template to FHIR Bundle entries.

The transformed *Observation* resource is added to the FHIR Bundle as an entry. The *GROWTH* section is transformed into multiple observations, as this section contains CDA *Observation* classes in <*component*> elements for different measurements taken during the procedure: weight, height, head circumference, fontanel measurements, and body mass index. The following reusable transformation components are used:

- *CdaGrowthSectionToFhirWeightObservation*
- *CdaGrowthSectionToFhirHeightObservation*
- *CdaGrowthSectionToFhirHeadCircumferenceObservation*
- *CdaGrowthSectionToFhirFontanelObservation*
- *CdaGrowthSectionToFhirBmiObservation*

The resulting Observation FHIR resources are added to the FHIR Bundle as entries. Referring to Figure 10 for the complexity of just the CDA document header component, we see how this approach encapsulates the complexity of a single document section and enhances clarity and high-level understanding of the clinical document's mapping to FHIR. From the data type level up to the CDA template level, the amount of code duplication is significantly reduced, as is the number of points of failure. At the same time, the clarity and comprehension of the transformations are greatly improved.

With the development of the "Notice of Growth" CDA to FHIR transformation, the following transformation components were created:

- *CdaClinicalDocumentHeaderToFhirBundle*
- *CdaAgeSectionToFhirObservation*
- *CdaGrowthSectionToFhirWeightObservation*
- *CdaGrowthSectionToFhirHeightObservation*
- *CdaGrowthSectionToFhirHeadCircumferenceObservation*
- *CdaGrowthSectionToFhirFontanelObservation*
- *CdaGrowthSectionToFhirBmiObservation*

Numerous transformation components have been created to convert CDA classes to FHIR resources and support the composition of section-level transformations. The essential components include the following:

- *CdaAssignedAuthorToFhirPractitioner*
- *CdaCustodianOrganizationToFhirOrganization*
- *CdaObservationToFhirObservation*
- *CdaOrganizationToFhirOrganization*
- *CdaPatientRoleToFhirPatient*
- *CdaEntryRelationshipToFhirObservationComponent*

The necessary data type transformation components include the following:

- *CdaAdToFhirExtendedContactDetail*
- *CdaCdToFhirCodeableConcept*
- *CdaCeToFhirCodeableConcept*
- *CdaIiToFhirIdentifier*
- *CdaIvlTsToFhirDateTime*
- *CdaPnToFhirHumanName*
- *CdaPqToFhirQuantity*
- *CdaRtoPqPqToFhirRatio*
- *CdaTelToFhirExtendedContactDetail*
- *CdaTsToFhirDate*

The *ConceptMap CdaAdministrativeGenderCodeToFhirGender* was also created. All these transformation components were designed to be reusable for the future development of transformation components from other CDA templates to FHIR bundles.

4 Analysis and discussions

4.1 Related work

This section provides a comprehensive review of the related work in the domain of data transformation, with a particular emphasis on the transformation process from CDA to FHIR. The related work can be systematically classified into three distinct categories: mapping languages, tools, and implementation projects. This categorization facilitates a more structured and in-depth analysis of the field.

4.1.1 Mapping languages

The concept of "Mapping Language" (or Data Transformation Language) lies in establishing a platform-independent specification that can be implemented across various programming languages (70). Model-to-model transformations are typically articulated in specialized domain-specific languages, often known as model transformation languages (MTLs) (71). MTLs encapsulate algorithms that delineate the process of converting elements from one model (or multiple models) into elements of another model (or multiple models). Declarative MTLs (DTLs) only provide logic constructs to express relations between elements in these candidate models, and the execution engine is responsible for synthesizing an execution plan that uses these relations to perform the model transformation.

Query/view/transformation: "Query/View/Transformation" (QVT) is a specification developed by the Object Management Group (OMG) to describe transformation rules between different data models in the Model-Driven Architecture (MDA) domain (72). The language was intended to support the declarative specification of model transformations, avoid imperative constructs, and support change propagation from one model to another as well as the bi- (or multi-) directional interpretation of transformations. However, its semantics have many unclear or unsatisfactory aspects that are not precisely defined in the standard (73). The QVT Core language (QVTc) uses pattern matching as the primary logic construct. Pattern matching is done over a flat set of variables by evaluating conditions over those variables against the candidate models (74).

eXtensible stylesheet language transformations: XSLT is a language used to transform XML documents into other document formats or other versions of XML.[2] XSLT is a powerful tool and a widely adopted language for transforming XML documents, including healthcare-related XML standards such as CDA. However, it is unsuitable for directly programming transformations of semantically complex models due to its low-level syntax (75). XSLT is also not a specialized language for medical data (76). One of its disadvantages is the mandatory use of XML language, which imposes limitations on use. It is also poorly readable, making it difficult to learn and debug (77).

Whistle: The Whistle Data Transformation Language provides a means to express mappings between schemes, enabling users to convert complex, nested data models into other equally complex and nested data formats (78). Whistle does not require a description of logical models for the data to be converted. The conversion requires only source data in JSON format and a map that describes the conversion rules. The result of the transformation is output data in JSON format.

Liquid templates: Liquid (79) is a templating language developed by Shopify that uses a combination of objects, tags, and filters inside template files to convert any JSON or XML format into another JSON format. A transformation engine is required to convert input data into output data based on a *.liquid template*. Microsoft FHIR Converter (80) is one such engine, processing Liquid templates to convert input data into validated FHIR format. It includes extended methods for FHIR data and is part of Microsoft's FHIR server implementation, available in the Microsoft Azure Health Data Services product (81). Users can upload custom templates to the Azure registry, which Azure Health Data Services can then use via an API endpoint for data transformation.

FHIR Mapping Language: The FHIR Mapping Language (FML) (40) is a relatively new QVT-based transformation language specifically designed to transform HL7 FHIR resources to/from alternative representations, including different logical data models, FHIR resources, C-CDA documents (42), etc. (82). FML is a part of the FHIR specification. Conceptually, FML is similar to XSLT:

(1) It consists of declarative rules that are automatically matched to input data
(2) It includes a sub-language (*FHIRPath*) to reference parts of source parse trees
(3) It can reference external functions written in different languages

The source input of FML supports any object models and rendering syntaxes that conform with OMG's Meta Object Facility (MOF)[3] language. MOF is a general formalism for representing object models as directed acyclic graphs (DAGs). MOF-compliant models can use various syntactic constructs to represent the classes, attributes, and attribute values of such graphs. The applications of this language encompass several scenarios:

- Mapping FHIR resources across different versions of FHIR
- Converting sections of HL7 C-CDA documents into multiple FHIR resources
- Translating HL7 V2 messages into multiple FHIR resources
- Adapting any structured data format into another structured data format, including mapping to multiple FHIR resources

[2] https://www.w3.org/TR/xslt-30.

[3] http://www.omg.org/mof/.

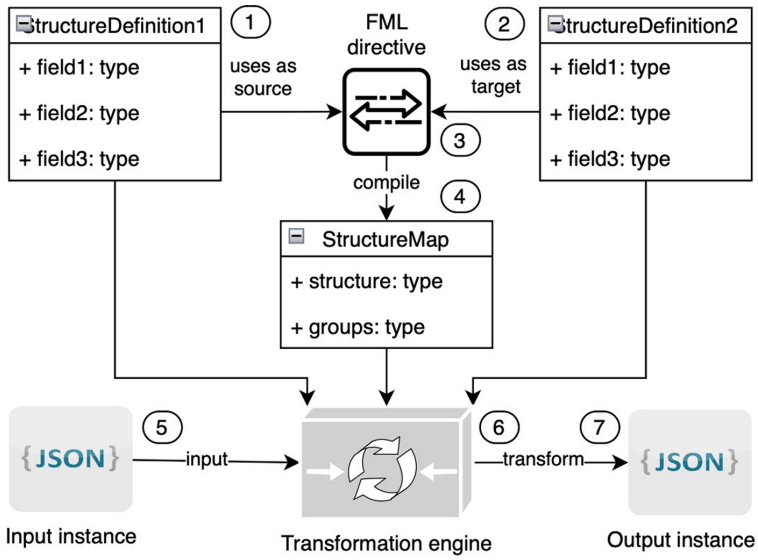

FIGURE 12
Components of FML transformation.

The technical specification of FML (40) has been published as an integral component of the FHIR specification (83). FML serves as a tool for transforming structured models from one form to another. Within the HL7 FHIR context, FML is utilized to map FHIR resources across different versions of FHIR. FML transformation requires the following (Figure 12):

- One input model (marked on the picture with the number "1")
- At least one output model (2)
- Human-readable transformation rules (also known as FML mapping directives) (3) that outline how to transform input into output
- A machine-processable transformation map (4) created as a result of the compilation transformation rules
- One input instance that corresponds to the input model in JSON or XML format (5)
- A transformation engine (6) that will transform the input instance to the output instance (7) based on models and transformation maps

4.1.2 Data transformation tools

NextGen connect: NextGen Connect (previously known as Mirth Connect) (84) is a robust, open-source healthcare integration engine widely used for its versatility and cost-effectiveness (85). One of its major strengths is its ability to support numerous data formats and protocols, such as HL7, XML, and JSON, making it highly adaptable to various healthcare systems (86). Its user-friendly interface and comprehensive documentation facilitate easier configuration and deployment, and the active community provides valuable support and resources. However, Mirth Connect has several drawbacks. Despite its user-friendly interface, it is primarily geared towards technical experts, making it challenging for domain experts without technical backgrounds to use it effectively (87). In our opinion, the learning curve is steep for new users unfamiliar with healthcare data standards and integration concepts. Performance can also be an issue with large-scale implementations, requiring careful optimization and resource management. Additionally, the clarity of implemented transformations can sometimes be lacking, making it difficult to understand and troubleshoot complex data flows (88). Furthermore, while the open-source version is feature-rich, some advanced features and enterprise-level support are only available in the paid version, which might limit its appeal to smaller organizations.

Other health data integration tools: Health data integration tools are essential for managing and transforming healthcare data, supporting interoperability within healthcare systems, and automating processes to realize cost savings. In addition to NextGen Connect, other well-known tools in this domain include Cloverleaf Integration Suite (89), Interfaceware Iguana (90), Corepoint Integration Engine (91), and Redox (92). Each tool offers numerous benefits, including connectivity and interface management, data transformation and workflow management, and support for various healthcare standards, protocols, and interfaces. They provide data mapping and support multiple data formats, leading to cost savings through reduced manual effort. However, there are challenges to consider when implementing these tools (93):

- *Complex implementation*: The process can be intricate, requiring IT professionals with expertise in healthcare data standards, protocols, and the specific tool's configuration.
- *Initial costs*: While cost savings can be realized in the long run, initial expenses associated with software licenses, hardware, and implementation can be challenging for smaller organizations.
- *Maintenance and support*: Regular updates, troubleshooting, and addressing issues are crucial for the tool's effectiveness, requiring dedicated resources.
- *Data mapping challenges*: Accurate and comprehensive data mapping can be challenging when dealing with disparate systems using different data standards and terminologies.
- *User training*: Staff may require training to use and navigate the tools effectively, and the learning curve can be costly.
- *Data security concerns*: Transmitting health data between systems raises data security concerns. Robust security measures are necessary to safeguard patient information and comply with data protection regulations.
- *Vendor lock-in*: Over-reliance on a specific tool or vendor can lead to potential issues if there are changes in the organization's strategy or the vendor's support changes.

FML implementations: The FHIR Mapping Language specification is implemented by code libraries such as the HAPI FHIR *StructureMap* implementation in Java (57) and its direct port to .Net (94), both of which offer transformation engines and open-source libraries. HAPI FHIR, a comprehensive Java library for FHIR, supports creating, parsing, and validating FHIR resources, providing robust tools for healthcare applications. The .Net FML implementation leverages these capabilities, bringing the same powerful functionality to the .Net ecosystem. Both libraries facilitate the transformation of healthcare data, ensuring interoperability and compliance with FHIR standards, which are crucial for modern healthcare systems.

Matchbox: Matchbox is an open-source initiative to support the testing and implementation of FHIR-based solutions (95). Matchbox utilizes the HAPI FHIR implementation, inheriting all its advantages while introducing additional flexibility for FML processing. Matchbox allows the preloading of FHIR implementation guides for conformance resources (*StructureMap*, *Questionnaire*, *CodeSystem*, *ValueSet*, *ConceptMap*, *NamingSystem*, *StructureDefinition*) and validates FHIR resources. Matchbox allows the defining of mapping in an FML text representation and its transformation into FHIR *StructureMap* resources. Matchbox applies the mapping to data to create FHIR-compatible data sets. Matchbox validates and executes FML transformations through the FHIR API, checking that the mapping conforms with the included validation stack.

4.1.3 Implementation projects

Austrian ELGA: The ELGA (Elektronische Gesundheitsakte) project launched in Austria is a nationwide EHR system designed to facilitate the exchange of medical documents across healthcare providers. ELGA uses CDA to manage medical data in a document-centric format. The project supports various document types, including Physician's Discharge Summaries, Nursing Discharge Summaries, Laboratory Reports, and Diagnostic Imaging Reports, with the addition of e-Medication reports covering prescription and medication summaries. To enhance interoperability and accessibility, recent efforts focus on mapping ELGA CDA documents to the FHIR standard using JSON mapping (96). Every element and section in JSON mapping has a "cda-path" that prescribes a rule for extracting data from a CDA document. This approach aims to generate International Patient Summaries (IPS) in FHIR format, enabling more granular access to health data and supporting cross-border healthcare data exchange within the European Union (26).

Italian patient summary: The Italian decree mandates that regional EHR systems support two types of documents: the Patient Summary and the Laboratory Report (27). The Patient Summary focuses on collecting the patient's most significant clinical information and uses the CDA format. During the *eHealthNet* project, a prototype was implemented for transforming the Patient Summary from CDA to FHIR. The proposed solution included the Mapping, Extractor, and Binding components. The Mapping component contains schemas defining correspondence between an element in FHIR and another in CDA. *XPath* was used for data extraction from CDA and binding to FHIR with a series of functions written in XSLT (27).

Swiss medications: The Swiss healthcare system has adopted the CDA standard, incorporating specific requirements unique to Switzerland (97). This has led to the creation of the CDA-CH standards (98). Switzerland transitioned to FHIR and developed equivalent FHIR-CH specifications for medication. To verify the equivalences, mappings have been defined with the FHIR mapping language, and Matchbox has been used for transformation from CDA to FHIR and back (99). To aid this transformation process, a consolidated library of CDA templates was employed (60). The use of FML in this context facilitates the automated transformation and validation of data, ensuring compliance with FHIR profiles and enhancing the utility of Swiss health data across various healthcare scenarios.

Estonian Andmevaatur: The *Andmevaatur* (Data Viewer) is a tool summarizing and visualizing patient data in the ENHIS (28). The ENHIS is built upon HL7 V3 and CDA standards (100). Due to the ever-increasing volume of documents, the task of gathering observations, procedures, vaccinations, and other clinical information from documents has become increasingly time-consuming for doctors (101). *Andmevaatur* uses *xQuery* to request CDA documents from the ENHIS database, transforms them into FHIR resources using a custom-developed mapping language, and forwards the resources to the user interface application for presentation. The custom-developed mapping language includes pairs of *XPath* and *FHIRPath* and a Java adapter for their execution. *XPath* is used for data extraction from CDA and *FHIRPath* is used for inserting data into the appropriate place in the FHIR resource. The development of an independent mapping language has been discontinued, and migration to FML is planned. Using *Andmevaatur*, doctors can save at least three minutes per visit, which is approximately 15 percent of the time typically spent interacting with a patient (101).

TABLE 2 Evaluation of artifacts

Artifact	Strict data models	Reuse	Native FHIR support	Execu- table software	Open-source	Visual editor
Query/View/Transformation (QVT) language (4.1.1)	+	+	−	−	+	−
Extensible Stylesheet Language Transformations (XSLT) (4.1.1)	+	+	−	−	+	−
Whistle (4.1.1)	−	+	−	+	+	−
Liquid (4.1.1)	−	+	−	+	+	−
FHIR Mapping Language (FML) (4.1.1)	+	+	+	−	+	−
FML implementations (4.1.2)	+	+	+	+	+	−
Integration tools (4.1.2)	+/−	+	−	+	−	+/−
Matchbox (4.1.2)	+	+	+	+	+	−
TermX (2.2)	+	+	+	+	+	+

Notes: "+" indicates that the criterion is met, while "−" indicates that it is not met.

4.2 Comparison of languages, tools, and implementations

To find the most suitable tool for our needs, we embarked on a comprehensive comparison of various languages, implementations, and tools. Our evaluation was based on a set of carefully developed criteria; the results are summarized in Table 2 and the conclusion is as follows:

- *Strict data model support*: DTL-based languages, such as FML, and their implementations provided robust support for strict data models.
- *Reuse of transformation*: We found that all languages used in evolution, along with their implementations and software, commendably support the reuse of transformations.
- *FHIR native support*: FML implementations, Matchbox, and TermX may be classified as tools with native FHIR support.
- *Executable software*: All implementations and software are classified as executable software.
- *Open-source license*: All languages, implementations, and software, except for NextGen Connect, and tools in the section "Other health data integration tools" are available under open-source licenses, promoting transparency and collaboration.
- *Visual transformation editor*: TermX and the health data integration tools stood out with their visual editors, which greatly facilitate the management of transformation flow.

After a comprehensive evaluation, it became evident that none of the existing implementations or tools were suitable, as they did not meet all of our selection criteria. This aligns with the health data interoperability issues highlighted in various recent papers by other implementers (27, 96).

In response to this, we developed the TermX FML Editor using the DS methodology. The designers behind TermX leveraged the existing FML language and the HAPI FHIR implementation, validating and reusing them to mitigate the risk of failure. Upon evaluating TermX, it was unequivocally clear that it was the only solution that met all of our selection criteria, thereby establishing it as the optimal choice for our needs.

4.3 Evaluation of visual reusable transformation rules

4.3.1 Toward federated interoperability in the EHDS

Ensuring federated interoperability (23, 24) is essential in the EHDS as it reduces administrative, operational, and international coordination costs. Federated systems store data in appropriate locations and formats, avoiding the complexity of large central repositories (102). This respects data sovereignty and privacy rules while allowing interoperability and independent innovation (103).

Centralized systems require significant infrastructure investment and management, which can be inefficient. Federated systems distribute these responsibilities, leveraging existing infrastructure and expertise and reducing compliance burdens with diverse regulatory frameworks. Federated semantic interoperability facilitates real-time data sharing, which is crucial for informed healthcare decision-making. By enabling seamless health data exchange, federated systems support innovative healthcare solutions, such as integrated care platforms and personalized medicine networks, enhancing care quality and patient outcomes.

Federated interoperability also supports EHDS initiative evaluations by providing a robust data integration and analysis framework, essential for assessing health interventions and informing policy decisions. Leveraging diverse data sources without extensive migration accelerates innovation and evaluation in healthcare. However, an effective system for semantic data transformation is required, as subsystems use different standards and models. The EHDS will inevitably need semantic data transformation, necessitating the evolution of user-friendly tools such as TermX.

4.3.2 Empowering domain experts

Achieving semantic interoperability is challenging due to the complexity of data transformation processes, which traditionally require significant technical expertise. The proposed techniques and TermX tool enable domain experts with minimal technical skills to participate effectively. The visual editor allows them to create and manage data transformation rules through an intuitive interface, democratizing the process and reducing reliance on technical specialists. This expedites development and deployment, improving the efficiency and scalability of interoperability initiatives.

The TermX tool explained in this paper allows domain experts to develop and validate data transformation rules, accommodating the evolving landscape of health standards and technologies (104). Direct involvement of domain experts ensures accuracy and relevance, as they bring a deep understanding of specific data and context. This collaboration fosters a more comprehensive approach to data transformation, enhancing the quality and reliability of interoperable data. The tool's validation features enable domain experts to test and refine transformation components, ensuring that transformed data meets expected standards and requirements and contributes to effective and trustworthy interoperability solutions.

4.3.3 Continuous adaptation to emerging innovations

Achieving federated semantic health data interoperability is crucial for supporting innovation within the EHDS (17). The healthcare data landscape constantly evolves, driven by innovations and new requirements. Semantic interoperability requires continuous adaptation. The proposed techniques and TermX tool support a flexible, modular approach to data transformation, adapting to new standards and technologies as they emerge. This ensures long-term interoperability and prevents obsolescence.

For instance, the transition from CDA to FHIR represents a significant shift in data structuring and exchange. As new versions of these standards are released, the tool must incorporate these changes, facilitating seamless data transformation. This capability allows healthcare organizations to leverage the latest advancements without significant disruptions or reengineering.

The evolving standards highlight the need for a collaborative approach to interoperability. The tool leverages collective expertise to stay updated with the latest developments by fostering a community-driven repository of transformation components and best practices. This promotes continuous improvement and innovation in health data interoperability.

4.3.4 Open FAIR access to routine clinical data

The FAIR (Findable, Accessible, Interoperable, Reusable) data principles are key enablers of secondary data use for societal benefit (105). Opening FAIR access to routine clinical data can drive advancements in medical research, clinical trials, public health, and policy-making (2–4, 106). Achieving FAIR access while maintaining privacy and security is challenging and requires robust technical solutions (18). Federated semantic interoperability offers a solution by keeping data in its original location, ensuring privacy, and enabling the integration and analysis of anonymized or pseudonymized data.

The proposed techniques and TermX tool support FAIR principles by providing a framework for transforming and integrating clinical data in a standardized manner. This ensures that data is findable and accessible, consistently represented, and understood. By facilitating data reuse through interoperable transformation rules, the tool enhances the utility of clinical data for secondary purposes. Leveraging routine clinical data for secondary use has profound societal implications, providing researchers with data for studies, enabling public health officials to monitor and respond to health threats, and guiding policymakers with evidence-based insight (107).

4.3.5 Integrating health data with other sectors

Health data is interconnected with data from sectors such as education, social services, the environment, and the economy (108, 109). Integrating health data with these sectors is essential for a holistic understanding of health determinants and outcomes, as the World Health Organization (WHO) recommends (110).

Although TermX was designed with FHIR support for health data interoperability, it is versatile enough to integrate and facilitate interoperability with other data sets beyond healthcare. This adaptability allows TermX to connect health data with various sectors, such as education, social services, the environment, and the economy. TermX supports a more comprehensive analysis of factors influencing health outcomes by enabling seamless data exchange across these domains. This flexibility ensures that TermX can serve as a powerful tool for creating holistic data ecosystems where health data is enriched by insights from other sectors, ultimately contributing to more informed decision-making and improved public health strategies.

4.3.6 Toward resolving three health data dilemmas

Klementi et al. (18) identified three health data dilemmas: accessibility, comprehensiveness, and ownership. The *accessibility dilemma* involves balancing health data access for improved outcomes with protecting sensitive information. Ensuring FAIR (Findable, Accessible, Interoperable, and Reusable) access often conflicts with data protection requirements (111–113). The *comprehensiveness dilemma* concerns creating a complete health record from fragmented data stored across various systems. Issues such as semantic interoperability and legal barriers impede the consolidation of data into a comprehensive personal health record (PHR) (114). The *ownership dilemma* addresses the conflict between individuals' rights to control their health data and the practical difficulties of exercising these rights (115, 116).

An EHDS architecture where individuals own and control their health data could use decentralized content-addressable storage networks (18). The proposed techniques and TermX tool create conditions that enable individuals to share their health data with healthcare professionals and ensure FAIR access to routine clinical data for secondary use (117, 118). This empowers more stakeholders to participate in the data transformation process, keeping health data interoperability at the forefront of healthcare innovation.

4.4 Implementation scenarios

4.4.1 Execution of the transformations in the single installation

The technical implementation of the solution encompasses both the design and transformation phases. This paper focuses on the design phase, wherein data models and transformations are developed. The resulting artifacts can be stored either in GitHub or on a FHIR server. The TermX Editor is utilized for

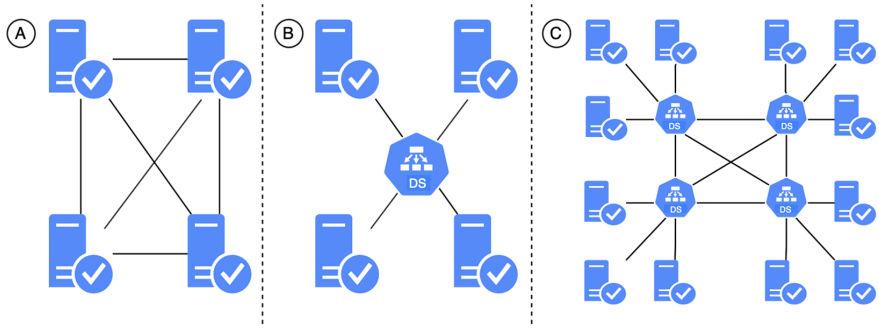

FIGURE 13
Possible topologies of EHDS: **(A)** every node is connected to every other node, **(B)** there is a central node to which all other nodes are connected, **(C)** a hybrid of topologies.

the design and testing of these transformations, but it is not required for their execution. For execution purposes, libraries such as HAPI FHIR, .Net, or their equivalents can be employed to compile and run the transformations. To enhance throughput, the application should support the caching of the utilized models (StructureDefinition instances) and compiled transformations (StructureMap instances). This application can function as a standalone service or as a module integrated into the FHIR server.

4.4.2 The transformations in the context of EHDS

When integrating two systems, two data models (source and target) and one set of transformations are required for one-way transformations or two sets for bidirectional transformations. If we consider that each medical system in the EHDS integrates with every other system and each has a unique data model, there will be N data models, resulting in an integration network with a complexity of $O(2^n)$ (Figure 13A). By creating a central model, we would have N+1 models and N (for one-way) or N*2 (for bidirectional) sets of transformations (Figure 13B). However, a single central model for all European countries is not realistic (9). It would be beneficial to reduce the number of models by creating smaller Data Spaces, where institutions within a country or region share a single model. Instead of a single central model, domain-specific Data Spaces could be established, connecting all EU laboratories (119), immunization records (120, 121), or radiology services into unified networks (Figure 13C). Such grouping would reduce the number of transformations and administrative burdens.

4.5 Limitations

4.5.1 Use-case-specific mapping of components

The current study was conducted and validated for a specific use case, namely the transformation of ENHIS documents. When comparing documents from Estonia with those from other countries, we find that documents of the same type, such as outpatient summaries, differ in the number of sections, section labeling, and terminology used. Additionally, country-specific extensions may be used. This implies that for each specific implementation, the representation at the business domain knowledge level may differ, and the set of transformations developed in this research study may require adaptation.

The foundational resources from the CDA and FHIR frameworks are highly compatible and could be suitable for use in any country. The ISO 23903 Interoperability and Integration Reference Architecture addresses the challenges associated with integrating such models and frameworks. Examples include mappings of HL7 V2 and HL7 V3 models and specifications, and the re-engineering and mapping of the higher-level specifications ISO 12967 Health Informatics Service Architecture and ISO 13940:2015 System of concepts to support continuity of care (122).

Although the detailing of base types in mapping may vary depending on the use case, for ENHIS, mapping of the CDA II to FHIR Identifier data types requires only the transformation of key attributes “root” to “system” and “extension” to “value” (Figure 4). However, in another information system, additional attributes such as “display” and “use” might be required, which we have not mapped, as this mapping is specific to the given use case. Nevertheless, it is easily generalizable if we extend the use case.

4.5.2 Mapping correctness

Actors from different scientific domains and disciplines, different communities, and different policy domains represent and understand related concepts differently (123). This decision on correct mapping is only possible at the business domain knowledge level, represented through domain ontologies and related terminologies.

- *Validation by analyst*. Business analysts, as domain experts, possess comprehensive knowledge of the domain’s ontology and terminology. They are responsible for planning and ensuring the accuracy of transformations. TermX is a robust tool specifically designed for analysts. Consequently, business

analysts are well-equipped to make transformation decisions and verify the accuracy of transformations by manually performing a reasonable number of tests.

- *Technical validation*. The technical validation of transformation correctness can be achieved through various methodologies. Section 2.3.4 elaborates on validation utilizing Natural Language Processing (NLP). Nevertheless, the ISO 23903 Interoperability and Integration Reference Architecture facilitates the accurate mapping of components across business, informational, computational, and engineering viewpoints. This framework supports the design and management of systems across diverse domains and contexts, thereby ensuring interoperability among ecosystem components (124).

Technical validation of transformations will make up future work.

5 Conclusion

Transforming health data from CDA to FHIR format is critical to achieving health data semantic interoperability. This paper presents generalized techniques for utilizing the TermX tool to develop reusable data transformation components and verify that the designed transformation components accurately transform data as expected. TermX leverages the FHIR Mapping Language to facilitate complex and technical data transformations. It is designed explicitly for domain experts, enabling them to develop and manage data transformation rules with minimal technical knowledge.

The pressing need for such a tool arises from the ongoing evolution of the ENHIS, which is transitioning EHRs from CDA to FHIR (22). This transition is not only a technical upgrade but also a strategic move to enhance health data's flexible and on-time semantic interoperability to improve the quality of clinical care and control healthcare costs, ensuring that patients' health information can be seamlessly shared and understood across systems and by healthcare practitioners in real time. Since vast amounts of historical EHR data in the ENHIS are stored in various HL7 CDA formats (15), transforming this data dynamically to FHIR as needed, rather than permanently, is essential. This approach utilizes federated semantic health data interoperability, ensuring that historical EHR data remains immutable but interoperable and accessible without requiring extensive and costly data migration efforts from one data repository and format to another.

The TermX tool was developed using the Design Science (DS) methodology, which emphasizes the creation and evaluation of artifacts designed to solve the problems identified. In the problem investigation phase, we conducted an analysis of languages, implementations, and tools to find a possible solution and tool to meet the ENHIS data transformation requirements. As we found no suitable solution or tool, and because the same health data interoperability issues were stressed in various recent papers, we developed TermX using the DS approach. TermX was designed (treatment design phase of DS) through the generalization, abstraction, and formalization of the needs of the ENHIS, ensuring that it is universal, usable, practical, and effective in most real-world health data transformation applications. The tool provides a visual editor for developing transformation components with FHIR Mapping Language support for transforming data from any data structure to any other. We evaluated (treatment validation phase of DS) that this tool might be usable and valuable for domain experts who may not have deep technical knowledge of information and communication technology. In the treatment implementation phase (not part of the DS but of the engineering cycle), we implemented the TermX solution with funding from the Estonian Business and Innovation Agency.

5.1 Research contribution

The primary business need addressed by the TermX tool is the efficient and validated transformation of health data from one data format to another. As healthcare organizations increasingly move toward adopting the FHIR standard, such tools are critical to bridge the semantic interoperability issues related to the concurrent utilization of legacy and new health data formats. Enabling domain experts to create and manage formal data transformation components in a simple WYSIWYG way using a visual editor, TermX reduces the need for technical specialists, which ultimately reduces costs and speeds up the deployment process needed to transform health data. Moreover, TermX ensures that data transformations can be carried out on the fly according to federated semantic interoperability, allowing data to be stored in different data formats while ensuring that healthcare providers have continuous and uniform access to both old and new data, in turn ensuring continuity of care and clinical decisions.

Socially, the implications of enhanced semantic interoperability are profound. Improved data interoperability means healthcare providers can share information more effectively, leading to better care coordination, reduced medical errors, and improved patient outcomes. This translates into more timely and accurate diagnoses, personalized treatment plans, and ultimately better patient health outcomes. Furthermore, integrating and analyzing data from diverse sources supports public health initiatives, research, and policy-making, contributing to the overall improvement of healthcare systems. The evaluation of the TermX tool demonstrated its effectiveness in developing reusable transformation components that domain experts can use for health data transformations. The tool was tested to ensure that the transformations were accurate and that they met the expected standards. The results showed that TermX could reliably perform the necessary transformations, supporting the hypothesis that a visual editor for the FHIR mapping language is both feasible and beneficial.

5.2 Future research and evaluation directions

While the TermX tool has shown promise, there are several areas for future research and development. One key area is the continuous improvement of the tool's user interface and experience, ensuring that it remains intuitive and accessible for domain experts. Additionally, expanding the tool's capabilities to handle more complex transformation scenarios and integrating machine learning techniques to suggest optimal transformation rules could further

enhance its utility. Another important direction is developing a comprehensive evaluation framework to continuously assess the quality and performance of the transformations. This framework could include metrics for measuring the accuracy, completeness, efficiency, user satisfaction, and adoption rates of transformations. Finally, fostering collaboration and knowledge-sharing among users of the TermX tool could lead to the development of a community-driven repository of transformation components and best practices. This repository could be a valuable resource for healthcare organizations worldwide, facilitating the broader adoption of FHIR and realizing truly interoperable health information systems.

5.3 Conclusion summary

In conclusion, the TermX tool represents a significant advancement in the quest for the unified federated semantic interoperability of health data. The tool addresses critical business and social needs by enabling domain experts to develop and manage transformation components with FHIR Mapping Language support. It supports the efficient and accurate transformation of health data, ensuring that historical data remains accessible and interoperable. As healthcare systems continue to evolve, tools such as TermX will play a crucial role in ensuring that data interoperability remains at the forefront of these advancements, ultimately leading to improved healthcare outcomes for patients and more efficient healthcare systems.

By addressing these critical areas, the TermX tool not only meets the immediate needs of the Estonian National Health Information System but also sets a precedent for other health systems seeking to enhance their data interoperability capabilities.

What was known on the topic:

(1) The EHDS aims to construct a health data-sharing ecosystem within the European Union, establishing rules and common standards to facilitate the use of EHRs.
(2) Each country that uses CDA tackles the transformation from CDA to FHIR in its own unique way, suggesting that there is no one-size-fits-all solution.
(3) Previously, no tools were available in the healthcare field for visualizing transformation with FHIR support.

What this study added to our knowledge:

(1) In the federated approach, systems that join the EHDS can store data in a location and format that suits them and transform the data to the EHDS standard in real time.
(2) TermX provides the ability to define and manage transformation components in a visual editor using the FML Mapping Language and strict data structures, such as FHIR resources and CDA classes.
(3) TermX enhances clarity, enables the reuse of transformation components, conceals the complexity of the FML mapping language, and allows analysts to quickly adapt to its usage.

Data availability statement

The TermX project[4] is available on GitHub, including the source code of TermX modules and applied projects. TermX modules include server, web application, and FML Editor (38). The source code of the developed CDA to FHIR transformations and the related presentations and screenshots are published in the TermX "cda2fhir" repository[5].

Author contributions

IB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; RR: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; GP: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing; PR: Funding acquisition, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work in the project "ICT programme" was supported by the European Union through the European Social Fund.

Acknowledgments

This design and evaluation of TermX FML Editor would not have been possible without the exceptional support of Jose Costa Teixeira, Oliver Egger, Vadim Peretokin, and Brian Postlethwaite.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

[4] https://github.com/termx-health.
[5] https://github.com/termx-health/cda2fhir.

References

1. Slawomirski L, Oderkirk J. Digital technology: making better use of health data. In: *New Health Technologies: Managing Access, Value and Sustainability* (2017). p. 185. doi: 10.1787/9789264266438-en

2. Wade TD. Refining gold from existing data. *Curr Opin Allergy Clin Immunol*. (2014) 14:181. doi: 10.1097/ACI.0000000000000051

3. PWC. *Transforming Healthcare Through Secondary use of Health Data*. Chicago: Benton Institute for Broadband & Society (2009). (accessed July 30, 2024).

4. Hackl WO, Ammenwerth E. SPIRIT: systematic planning of intelligent reuse of integrated clinical routine data: a conceptual best-practice framework and procedure model. *Methods Inf Med*. (2016) 55:114–24. doi: 10.3414/ME15-01-0045

5. European Commission. European Health Data Space. Available online at: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en (accessed June 29, 2024).

6. Hussein R, Scherdel L, Nicolet F, Martin-Sanchez F. Towards the European health data space (EHDS) ecosystem: a survey research on future health data scenarios. *Int J Med Inf*. (2023) 170:104949. doi: 10.1016/j.ijmedinf.2022.104949

7. European Commission. Questions and answers—European Health Data Space (EHDS). Available online at: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_22_2712/QANDA_22_2712_EN.pdf (accessed July 30, 2024).

8. van de Gronden J, Veenbrink M. EHDS And free movement of patients: what EU intervention is needed? *Eur J Health Law*. (2024) 31:249–84. doi: 10.1163/15718093-bja10125

9. Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I, et al. The European health data space: too big to succeed? *Health Policy*. (2023) 135:104861. doi: 10.1016/j.healthpol.2023.104861

10. Lalova-Spinks T, Saesen R, Silva M, Geissler J, Shakhnenko I, Camaradou JC, et al. Patients' knowledge, preferences, and perspectives about data protection and data control: an exploratory survey. *Front Pharmacol*. (2024) 14:1280173. doi: 10.3389/fphar.2023.1280173

11. Schmitt T, Cosgrove S, Pajić V, Papadopoulos K, Gille F. What does it take to create a European health data space? International commitments and national realities. *Z Evid Fortbild Qual Gesundhwes*. (2023) 179:1–7. doi: 10.1016/j.zefq.2023.03.011

12. Quinn P, Ellyne E, Yao C. Will the GDPR restrain health data access bodies under the European health data space (EHDS)? *Comput Law Secur Rev*. (2024) 54:105993. doi: 10.1016/j.clsr.2024.105993

13. Sharon T, Gellert R. Regulating big tech expansionism? Sphere transgressions and the limits of Europe's digital regulatory strategy. *Inform Commun Soc*. (2023):1–18. doi: 10.1080/1369118X.2023.2246526

14. van Kessel R, Wong BLH, Forman R, Gabrani J, Mossialos E. The European health data space fails to bridge digital divides. *Br Med J*. (2022) 378:1–2.

15. Metsallik J, Ross P, Draheim D, Piho G. Ten years of the e-health system in Estonia. In: Rutle A, Lamo Y, MacCaull W, Iovino L, editors. *CEUR Workshop Proceedings [3rd International Workshop on (Meta) Modelling for Healthcare Systems (MMHS)]*. vol. 2336 (2018). p. 6–15.

16. ELGA. Technischer aufbau im Überblick. Available online at: https://www.elga.gv.at/technischer-hintergrund/technischer-aufbau-im-ueberblick/ (accessed June 29, 2024)

17. Raab R, Küderle A, Zakreuskaya A, Stern AD, Klucken J, Kaissis G, et al. Federated electronic health records for the European health data space. *Lancet Digit Health*. (2023) 5:e840–7. doi: 10.1016/S2589-7500(23)00156-5

18. Klementi T, Piho G, Ross P. A reference architecture for personal health data spaces using decentralized content-addressable storage networks. *Front Med (Lausanne)*. (2024) 11:1–24. doi: 10.3389/fmed.2024.1411013

19. de Mello BH, Rigo SJ, da Costa CA, da Rosa Righi R, Donida B, Bez MR, et al. Semantic interoperability in health records standards: a systematic literature review. *Health Technol (Berl)*. (2022) 12:255–72. doi: 10.1007/s12553-022-00639-w

20. Chatterjee A, Pahari N, Prinz A. Hl7 fhir with snomed-ct to achieve semantic and structural interoperability in personal health data: a proof-of-concept study. *Sensors*. (2022) 22:3756. doi: 10.3390/s22103756

21. Tsinale HL, Mbugua S, Luvanda A. Architectural health data standards and semantic interoperability: a comprehensive review in the context of integrating medical data into big data analytics. *Int J Eng Appl Sci Technol*. (2023) 8:17–30. doi: 10.33564/IJEAST.2023.v08i04.002

22. Bossenko I, Piho G, Ross P. *Modelling a Patient Identifier System in the Estonian National Health Information System*. vol. 531. Cham: Springer (2024). p. 14.

23. Ciampi M, De Pietro G, Esposito C, Sicuranza M, Donzelli P. A federated interoperability architecture for health information systems. *Int J Internet Protoc Technol*. (2013) 7:189–202. doi: 10.1504/IJIPT.2013.058646

24. Tu Z, Zacharewicz G, Chen D. A federated approach to develop enterprise interoperability. *J Intell Manuf*. (2016) 27:11–31. doi: 10.1007/s10845-013-0868-1

25. Smits M, Kramer E, Harthoorn M, Cornet R. A comparison of two detailed clinical model representations: FHIR and CDA. *Eur J Biomed Inform*. (2015) 11:7–17. doi: 10.24105/ejbi.2015.11.2.3

26. Dimitrov A, Duftschmid G. Generation of FHIR-based international patient summaries from ELGA data. (2022) 293:1–8. doi: 10.3233/SHTI220339

27. Mercorella M, Ciampi M, Esposito M, Esposito A, De Pietro G. An architectural model for extracting FHIR resources from CDA documents. *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (IEEE)* (2016). p. 597–603.

28. Bertl M, Kankainen KJI, Piho G, Draheim D, Ross P. Evaluation of data quality in the Estonia national health information system for digital decision support. *Proceedings of the 3rd International Health Data Workshop* (2023). p. 1–13.

29. IPS Web Editorial Team. The International Patient Summary. (2024). Available online at: https://international-patient-summary.net/ (accessed July 30, 2024).

30. Egger O, Ziegler M, Eastwood R. Implementation Guide CDA FHIR Maps. (2024). Available online at: https://github.com/hl7ch/cda-fhir-maps/ (accessed July 30, 2024).

31. Arlow J, Neustadt I. *Enterprise Patterns and MDA: Building Better Software with Archetype Patterns and UML*. Boston: Addison-Wesley Professional (2004).

32. Wieringa RJ. *Design Science Methodology for Information Systems and Software Engineering*. Springer (2014).

33. Ivanova M, Bossenko I, Piho G. The comparison of the terminology servers. *Lect Notes Bus Inf Process*. (2024) 531:12.

34. Bossenko I, Piho G, Ross P. Termx: a game changer in the healthcare interoperability. *Stud Health Technol Inform*. (2024) 316:88–9. doi: 10.3233/SHTI240352

35. TermX. TermX homepage. (2024). Available online at: https://termx.org/ (accessed July 30, 2024).

36. Ivanova M, Bossenko I, Piho G. TermX - bridging the gap: implementing CTS2 and FHIR compatible terminology server. *International Congress on Information and Communication Technology (Springer), 10. is Accepted by ICICT 2024 (9th International Congress on Information and Communication Technology)*; London, UK (2024).

37. Marquis M, Bossenko I, Ross P. Radlex and SNOMED CT integration: a pilot study for standardizing radiology classification. *Insights Imaging*. (2025). in press.

38. Bossenko I, Piho G, Ivanova M, Ross P. Termx: the semantic interoperability, knowledge management and sharing platform. *SoftwareX*. (2024) 27:101839. doi: 10.1016/j.softx.2024.101839

39. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. *MIS Q*. (2004) 28:75–105. doi: 10.2307/25148625

40. HL7. FHIR Mapping Language. (2023). Available online at: http://hl7.org/fhir/mapping-language.html (accessed July 30, 2024).

41. HL7. FHIR StructureMap. (2024). Available online at: http://hl7.org/fhir/structuremap.html (accessed July 30, 2024).

42. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 clinical document architecture, release 2. *J Am Med Inform Assoc*. (2006) 13:30–9. doi: 10.1197/jamia.M1888

43. HL7. FHIR is a standard for health care data exchange, published by HL7®. (2022). Available online at: http://hl7.org/fhir/ (accessed July 30, 2024).

44. Benson T, Grieve G. *Principles of Health Interoperability*. Cham: Springer International (2021). p. 21–40.

45. Braunstein ML, Braunstein ML. *Health Informatics on FHIR: How HL7's API is Transforming Healthcare*. Cham: Springer (2022).

46. Tanjga N, Krauss O. Dual implementation guides in FHIR and CDA. *Stud Health Technol Inform*. (2024) 313:49–54.

47. Eggebraaten TJ, Tenner JW, Dubbels JC. A health-care data model based on the HL7 reference information model. *IBM Syst J*. (2007) 46:5–18. doi: 10.1147/sj.461.0005

48. HL7. FHIR and Architectural Principles. (2024). Available online at: http://hl7.org/fhir/overview-arch.html#principles (accessed July 30, 2024).

49. Kramer MA. Reducing FHIR "profiliferation": a data-driven approach. *AMIA Annu Symp Proc*. (2022) 2022:634.

50. HL7. CDA Interoperability Considerations. (2024). Available online at: https://www.hl7.org/implement/standards/fhir/comparison-cda.html#7.18.3.2 (accessed July 30, 2024).

51. HL7. Standards. Clinical Document Architecture (CDA®). (2024). Available online at: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=492 (accessed July 30, 2024).

52. HL7. FHIR. Change Management and Versioning. (2024). Available online at: https://www.hl7.org/fhir/versions.html (accessed July 30, 2024).

53. Bossenko I, Piho G, Ross P. Forward and backward compatibility design techniques applying the HL7 FHIR standard. *HEDA-2022: The International Health Data Workshop*; June 19–24, 2022; Bergen (Norway: CEURAT) (2022). p. 14.

54. HL7. FHIR Structure Definition. (2024). Available online at: http://hl7.org/fhir/structuredefinition.html (accessed July 30, 2024).

55. FHIREST Community. FHIR Server. (2023). Available online at: https://github.com/fhirest (accessed July 30, 2024).

56. Smile Digital Health. HAPI FHIR. (2024). Available online at: https://hapifhir.io/ (accessed July 30, 2024).

57. HL7. HAPI FHIR Structure Map Utilities. (2023). Available online at: https://github.com/hapifhir/org.hl7.fhir.core/blob/master/org.hl7.fhir.r5/src/main/java/org/hl7/fhir/r5/utils/structuremap/StructureMapUtilities.java (accessed July 30, 2024).

58. HL7. FHIR Datatypes. (2024). Available online at: http://hl7.org/fhir/datatypes.html (accessed July 30, 2024).

59. HL7. Bundle. (2024). Available online at: http://hl7.org/fhir/bundle.html (accessed July 30, 2024).

60. HL7. FHIR Definition of Clinical Document Architecture. (2024). Available online at: https://build.fhir.org/ig/HL7/CDA-core-sd/ (accessed July 30, 2024).

61. Õitspuu R. An Analysis of the Data Composition of the Estonian National Health Information System and Recommendations for the Implementation of the HL7 FHIR standard. (2024). Available online at: https://digikogu.taltech.ee/et/Item/6b137f0a-efda-42b1-9653-d593d317f712 (accessed July 30, 2024).

62. Kandel S, Paepcke A, Hellerstein J, Heer J. Wrangler: interactive visual specification of data transformation scripts. *Proceedings of the Sigchi Conference on Human Factors in Computing Systems*. (2011). p. 3363–72. doi: 10.1145/1978942.1979444

63. European Commission. About Technology Readiness Levels. (2024). Available online at: https://euraxess.ec.europa.eu/career-development/researchers/ (accessed July 30, 2024).

64. Jackups R Jr. Fda regulation of laboratory clinical decision support software: is it a medical device? *Clin Chem*. (2023) 69:327–9. doi: 10.1093/clinchem/hvad011

65. Weissman GE. Fda regulation of predictive clinical decision-support tools: what does it mean for hospitals? *J Hosp Med*. (2021) 16:244. doi: 10.12788/jhm.3450

66. Ontañón S. An overview of distance and similarity functions for structured data. *Artif Intell Rev*. (2020) 53:5309–51. doi: 10.1007/s10462-020-09821-w

67. Ramos J. Using tf-idf to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning (Citeseer)*. vol. 242 (2003). p. 29–48.

68. HL7. ConceptMap. (2024). Available online at: http://hl7.org/fhir/conceptmap.html (accessed July 30, 2024).

69. HL7. FHIR Shorthand. (2024). Available online at: https://hl7.org/fhir/uv/shorthand (accessed July 30, 2024).

70. Sibley EH, Taylor RW. A data definition and mapping language. *Commun ACM*. (1973) 16:750–9. doi: 10.1145/362552.362555

71. Rodriguez HH, Kolovos DS. Declarative model transformation execution planning. *OCL@ MoDELS (CEUR-WS)*. vol. 1756 (2016). p. 105–20.

72. Kurtev I. State of the art of QVT: a model transformation language standard. *Applications of Graph Transformations with Industrial Relevance: Third International Symposium, AGTIVE 2007, Kassel, Germany, October 10–12, 2007, Revised Selected and Invited Papers 3*. Kassel: Springer (2008). p. 377–93.

73. Lano K, Kolahdouz-Rahimi S. Implementing QVT-R via semantic interpretation in UML-RSDS. *Softw Syst Mod*. (2021) 20:725–66. doi: 10.1007/s10270-020-00824-3

74. Object Management Group (OMG). Meta Object Facility (MOF) 2.0 Query/View/ - Transformation Specification. Version 1.3. (2016). Available online at: https://www.omg.org/spec/QVT/About-QVT (accessed July 30, 2024).

75. Li D, Li X, Stolz V. QVT-based model transformation using XSLT. *ACM SIGSOFT Softw Eng Notes*. (2011) 36:1–8. doi: 10.1145/1921532.1921563

76. Seol YH, Johnson SB, Starren J. Use of the extensible stylesheet language (xsl) for medical data transformation. *Proceedings of the AMIA Symposium*. Washington: American Medical Informatics Association (1999). p. 142.

77. Bonifati A. *Reactive Services for XML Repositories*. [Ph.D. Thesis, Citeseer]. (2002).

78. Google Cloud Platform. Whistle Data Transformation Language. (2024). Available online at: https://github.com/GoogleCloudPlatform/healthcare-data-harmonization (accessed July 30, 2024).

79. Shopify. Liquid template language. (2024). Available online at: https://shopify.github.io/liquid (accessed July 30, 2024).

80. Microsoft. FHIR Converter. (2024). Available online at: https://github.com/microsoft/FHIR-Converter (accessed July 30, 2024).

81. Microsoft. Azure Health Data Services. (2024). Available online at: https://azure.microsoft.com/en-us/products/health-data-services (accessed July 30, 2024).

82. HL7. HL7 CIMI Logical Model: Analysis Normal Form (ANF), Release 1. (2020). Available online at: https://confluence.hl7.org/display/ADDC/HL7+CIMI+Logical+Model%3A+Analysis+Normal+Form+%28ANF%29%2C+Release+1 (accessed July 30, 2024).

83. HL7. FHIR documentation. (2024). Available online at: http://hl7.org/fhir/documentation.html (accessed July 30, 2024).

84. NextGen Healthcare. Mirth Connect. (2006). Available online at: https://www.nextgen.com/solutions/interoperability/mirth-integration-engine (accessed July 30, 2024).

85. Castanheira A, Peixoto H, Machado J. Overcoming challenges in healthcare interoperability regulatory compliance. *Ambient Intelligence–Software and Applications: 11th International Symposium on Ambient Intelligence*. Boca Raton: Productivity Press (2021). p. 44–53.

86. Alkarkoukly S, Kamal MM, Beyan O. Breaking barriers for interoperability: a reference implementation of csv-fhir transformation using open-source tools. In: Hägglund M, Blusi M, Bonacina S, Nilsson L, Madsen IC, Pelayo S, et al. editors. *Caring is Sharing–Exploiting the Value in Data for Health and Innovation*. vol. 302. IOS Press (2023). p. 43–7.

87. Randmaa R, Bossenko I, Klementi T, Piho G, Ross P. Evaluating business meta-models for semantic interoperability with fhir resources. *HEDA-2022: The International Health Data Workshop*; June 19–24, 2022; Bergen (Norway: CEURAT) (2022). p. 14.

88. Malowidzki M, Południowe P, Kozak M, Bereziński P, Piotrowski R. Integrated solutions for NATO forward medical evacuation: experiences and insights. *NATO Science and Technology Organization Symposium (ICMCIS)*; 16–17 May 2023; Skopje, North Macedonia (2023). p. 1–6.

89. Infor. Cloverleaf Integration Suite. (2023). Available online at: https://www.infor.com/products/cloverleaf (accessed July 30, 2024).

90. Interfaceware. Iguana - HL7 Interface Engine and Healthcare Integration. (2023). Available online at: https://www.interfaceware.com/iguana (accessed July 30, 2024).

91. Rhapsody Health. Corepoint Integration Engine. (2023). Available online at: https://rhapsody.health/solutions/corepoint/ (accessed July 30, 2024).

92. Redox Engine. Redox Healthcare Integration and Interoperability. (2023). Available online at: https://www.redoxengine.com/ (accessed July 30, 2024).

93. Arzt N, Berry M. Open source and public health: a primer. *Med Res Arch*. (2019) 7:1–11. doi: 10.18103/mra.v7i12.2011

94. Postlethwaite B. Direct port of the Java implementation of the HL7 FHIR Mapping Language on the Microsoft. NET. (2023). Available online at: https://github.com/brianpos/fhir-net-mappinglanguage (accessed July 30, 2024).

95. Ahdis. Matchbox. (2024). Available online at: https://matchbox.health/ (accessed July 30, 2024).

96. Rinner C, Duftschmid G. *Bridging the gap Between HL7 CDA and HL7 FHIR: A JSON Based Mapping*. vol. 223. IOS Press (2016). p. 100–6.

97. Denecke K, Dittli PA, Kanagarasa N, Nüssli S. Facilitating the information exchange using a modular electronic discharge summary. *Health Inform Meets EHealth*. (2018) 248:72–9. doi: 10.3233/978-1-61499-858-7-72

98. HL7 Switzerland. Swiss Implementation Guides. (2024). Available online at: https://www.fhir.ch/ (accessed July 01, 2024).

99. HL7 Switzerland. CDA-FHIR-Maps. (2024). Available online at: https://fhir.ch/ig/cda-fhir-maps (accessed July 30, 2024).

100. Sepper R, Ross P, Tiik M. Nationwide health data management system: a novel approach for integrating biomarker measurements with comprehensive health records in large populations studies. *J Proteome Res*. (2011) 10:97–100. doi: 10.1021/pr1007784

101. Estonian Health and Welfare Information Systems Centre. Andmevaatur (in Estonian). (2024). Available online at: https://www.tehik.ee/uudis/meditsiinitootajatele-loodi-uus-lahendus-terviseandmetega-tutvumiseks (accessed July 30, 2024).

102. Alsamhi SH, Myrzashova R, Hawbani A, Kumar S, Srivastava S, Zhao L, et al. Federated learning meets blockchain in decentralized data-sharing: healthcare use case. *IEEE Internet Things J*. (2024) 11:19602–15. doi: 10.1109/JIOT.2024.3367249

103. Fahey BA. Data federalism. *Harv Law Rev*. (2021) 135:1007.

104. Carroll WM. *Digital Health and new Technologies. Nursing and Informatics for the 21st Century-Embracing a Digital World, Book 1*. Productivity Press (2022). p. 29–48.

105. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: scoping review. *J Med Internet Res*. (2023) 25:e45013. doi: 10.2196/45013

106. Hughes N, Kalra D. *Data Standards and Platform Interoperability. Real-World Evidence in Medical Product Development*. Cham: Springer (2023). p. 79–107. doi: 10.1007/978-3-031-26328-6_6

107. Lee JS, Tyler AR, Veinot TC, Yakel E. Now is the time to strengthen government-academic data infrastructures to jump-start future public health crisis response. *JMIR Public Health Surv*. (2024) 10:e51880. doi: 10.2196/51880

108. Sadana R, Harper S. Data systems linking social determinants of health with health outcomes: advancing public goods to support research and evidence-based policy and programs. *Public Health Rep*. (2011) 126:6–13. doi: 10.1177/00333549111260S302

109. Dally D, Amith M, Mauldin RL, Thomas L, Dang Y, Tao C. A semantic approach to describe social and economic characteristics that impact health outcomes (social determinants of health): ontology development study. *Online J Public Health Inform*. (2024) 16:e52845. doi: 10.2196/52845

110. Marmot M, Friel S, Bell R, Houweling TA, Taylor S; Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet*. (2008) 372(9650):1661–9. doi: 10.1016/S0140-6736(08)61690-6

111. Determann L. Healthy data protection. *Mich Technol Law Rev*. (2019) 26:229.

112. Demotes-Mainard J, Cornu C, Guerin A, Bertoye PH, Boidin R, Bureau S, et al. How the new European data protection regulation affects clinical research and recommendations? *Therapies*. (2019) 74:31–42. doi: 10.1016/j.therap.2018.12.004

113. McKay F, Treanor D, Hallowell N. Inalienable data: ethical imaginaries of de-identified health data ownership. *SSM Qual Res Health*. (2023) 4:100321. doi: 10.1016/j.ssmqr.2023.100321

114. De Zegher I, Norak K, Steiger D, Müller H, Kalra D, Scheenstra B, et al. Artificial intelligence based data curation: enabling a patient-centric European health data space. *Front Med (Lausanne)*. (2024) 11:1365501. doi: 10.3389/fmed.2024.1365501

115. Ballantyne A. How should we think about clinical data ownership? *J Med Ethics*. (2020) 46:289–94. doi: 10.1136/medethics-2018-105340

116. Martani A, Geneviève LD, Elger B, Wangmo T. It's not something you can take in your hands. Swiss experts' perspectives on health data ownership: an interview-based study. *BMJ Open*. (2021) 11:e045717. doi: 10.1136/bmjopen-2020-045717

117. Ma L, Liao Y, Fan H, Zheng X, Zhao J, Xiao Z, et al. Phdmf: a flexible and scalable personal health data management framework based on blockchain technology. *Front Genet*. (2022) 13:877870. doi: 10.3389/fgene.2022.877870

118. Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer ML, Deserno T, Chronaki C, et al. From raw data to FAIR data: the FAIRification workflow for health research. *Methods Inf Med*. (2020) 59:e21–32. doi: 10.1055/s-0040-1713684

119. HL7 Europe. Implementation Guide "HL7 Europe Laboratory Report". (2024). Available online at: https://build.fhir.org/ig/hl7-eu/laboratory/ (accessed July 30, 2024).

120. WHO. WHO Immunization Implementation Guide. (2024). Available online at: https://worldhealthorganization.github.io/smart-immunizations/ (accessed July 30, 2024).

121. WHO. GAVI, The Vaccine Alliance. (2024). Available online at: https://www.who.int/europe/about-us/partnerships/global-health-partnerships/gavi-alliance (accessed July 30, 2024).

122. Blobel B, Ruotsalainen P, Oemig F, Giacomini M, Sottile PA, Endsleff F. Principles and standards for designing and managing integrable and interoperable transformed health ecosystems. *J Pers Med*. (2023) 13:1579. doi: 10.3390/jpm13111579

123. Blobel B, Kalra D. Managing healthcare transformation towards p5 medicine. *Front Med (Lausanne)*. (2023) 10:1244100. doi: 10.3389/fmed.2023.1244100

124. Blobel B, Oemig F, Ruotsalainen P, Brochhausen M, Giacomini M. The representational challenge for designing and managing 5p medicine ecosystems. *Stud Health Technol Inform*. (2024) 314:3–13. doi: 10.3233/SHTI240047

Appendix 3

III

I. Bossenko, K. Linna, G. Piho, and P. Ross, “Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia,” in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pp. 882–885, 2023

Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in the Infectious Disease Information System of Estonia

Igor Bossenko
Department of Health Technologies
TalTech, Tallinn, Estonia
igor.bossenko@taltech.ee

Kerli Linna
Department of Health Technologies
TalTech, Tallinn, Estonia
kerli.linna@taltech.ee

Gunnar Piho
Department of Software Science
TalTech, Tallinn, Estonia
gunnar.piho@taltech.ee

Peeter Ross
Department of Health Technologies
TalTech, Tallinn, Estonia
peeter.ross@taltech.ee

ABSTRACT

Objective: Many countries have successfully implemented HL7 V3 and CDA (Clinical Document Architecture) standards to ensure document-based interoperability between EHRs (Electronic Health Records), registries, and healthcare institutions [26] [32] [5]. The biggest drawback of the HL7 CDA document-based approach is the timing of sharing the information. The document is generally shared once all the agreed data elements have been precisely filled in and the necessary confirmations received. Today, Estonia is transitioning the Estonian National Health Information System (ENHIS) from a document-based approach to an event-based approach by utilising the HL7 FHIR (Fast Healthcare Interoperability Resources) standard. During this transition, one of the tasks is to describe the patient's socioeconomic status according to the FHIR profile.

Method: As part of the project, the CDA-based notification of infectious patients, including patients' socioeconomic status, in the Infectious Disease Information System (NAKIS - the acronym of the system in Estonian) of Estonia is analysed. This analysis pays special attention to education and employment data, which is currently part of patients' socioeconomic status. The existing employment and education-related FHIR profiles, openEHR clinical models and SNOMED and LOINC terminology are studied to find appropriate specifications for patients' socioeconomic status.

Results: As a result of the project, similar and/or suitable FHIR resources and terminology will be chosen. Structural changes will be made in the data collection process, where several simple values will be selected instead of one complex value. New FHIR profiles will be created, modelling new FHIR-based data structures and describing the terminology.

Summary: It is possible to reuse information from existing official sources rather than collect it from patients. Socioeconomic status is a temporal representation of the patient's education, employment and other similar information about the patient. We propose a universal and reusable methodology for migrating HL7 CDA documents and their components into FHIR resources.

SAC'23, March 27 –April 2, 2023, Tallinn, Estonia

ACM ISBN 978-1-4503-9517-5/23/03…$15.00
https://doi.org/10.1145/3555776.3577836

CCS CONCEPTS

• **Information systems** → **Enterprise applications**; • **Software and its engineering** → **Software architectures**; **Model-driven software engineering**;

KEYWORDS

HL7 Fast Health Interoperability Resources (HL7 FHIR), Estonian National Health Information System (ENHIS), electronic health record (EHR), interoperability, socioeconomic status, education level, occupation, HL7 Clinical Document Architecture (HL7 CDA), openEHR, SNOMED, LOINC

ACM Reference Format:
Igor Bossenko, Kerli Linna, Gunnar Piho, and Peeter Ross. 2023. Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in the Infectious Disease Information System of Estonia. In *Proceedings of ACM SAC Conference (SAC'23).* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3555776.3577836

1 PROBLEM

The Estonian National Health Information System (ENHIS) is a set of healthcare services that cover many aspects of healthcare - discharge summaries, referrals, e-prescriptions, the national appointments system, etc. [26]. ENHIS development began in 2005 and is based on HL7 V3 and CDA standards. HL7 CDA is a document markup standard that specifies the structure and semantics of clinical documents [2]. One of the data exchange services offered by ENHIS is NAKIS (Nakkushaiguste infosüsteem (in Estonian), Infectious Diseases Information System) [40], which serves the purpose of sharing information about infectious diseases and suspected infections (including AIDS, hepatitis and COVID-19) through the NAKIS to the register of infectious diseases. Regarding the COVID-19 pandemic, the NAKIS notice was one of the most frequently used notices from 2020 to 2021. To better analyse the subjects of infections, NAKIS requires the collection of social characteristics like occupation, employment organisation and educational or

preschool institution. Socioeconomic status (SES) is an attribute of a person's economic and social status that tends to be associated with a person's general state of health. [1]. In the ENHIS, this attribute currently combines three different information of socioeconomic status: education, income and occupation. According to the Estonian digital health master plan, ENHIS will be migrated from HL7 V3 and CDA to FHIR. HL7 FHIR [12] is an international healthcare information exchange standard that provides a range of predefined resources, the possibility to expand these resources and a framework for exchanging these resources between interested parties.

Objectives

This paper analyses the following questions. What resources are suitable for submitting socioeconomic status? What similar FHIR profiles and terminologies exist in the world? How to create FHIR profiles and terminology suitable for transmitting socioeconomic status [3]?

2 RELATED WORKS

In recent years, several health-related clinical data models have been proposed. The most popular clinical data models with technological standards are HL7 V2, HL7 CDA and HL7 FHIR. Several groups have tried to create mappings between standards or create technology-agnostic clinical models and mappings through these models. Smits et al. use a Detailed Clinical Model (DCM), and XSLT transformations between CDA and FHIR [34]. The Lantana Group provides tools for CDA transformation to FHIR [4]. In the Austrian EHR, Rinner and Duftschmid use a combination of XPATH and custom JSON mapping for conversion from CDA to FHIR [33]. Mercorella uses XPATH and custom XSLT mappings from CDA to FHIR in the Italian Patient Summary project [25]. Consolidated CDA (CCDA) on FHIR [8] conversion support was developed on top of the Argonaut Project to fully map CDA documents based on standard templates. The researchers point out that a universal approach leads to possible data loss, changes in meaning and lack of inter-convertibility. Solutions with custom mappings (Austrian, Italian) are more clinically precise. The situation where the original CDA resource must be divided into FHIR resources depending on the content can be solved with the help of custom mapping.

3 SOLUTION

It is a common opinion that characteristics reflecting the patient's current status, such as socioeconomic status, must be part of the Patient resource. The FHIR specification states, "The data in the Resource covers the 'who' information about the patient: its attributes are focused on the demographic information necessary to support the administrative, financial and logistic procedures" [13]. FHIR allows the addition of new elements as an extension [11]. For example, US Core Patient [19] includes information about race and ethnicity. According to the informational system modelling principle element, it is reasonable to add an attribute to a resource if it reflects an event that is complete, i.e. it changes very rarely, such as 1-2 times during the life of the resource [27]. If the resource changes more frequently, it should be recorded as an observable with a measurement date and/or period. Another factor to consider when designing a resource is knowing the history of the attribute. If the history is not essential in the current moment and will not become important in the future, then the attribute can be added as a resource attribute; otherwise, it must be observable by itself. Socioeconomic status may change several times a year, and multiple indicators may be simultaneously valid. Also, a historical overview of changes in the patient's socioeconomic status may be clinically valuable. Therefore, socioeconomic status cannot be an attribute or an extension of the Patient resource but has to be a separate FHIR Observation resource with its profile or profiles.

4 METHODOLOGY

Based on the principles explained in the previous sections, a list of keywords was compiled: education, school, work, workplace, occupation, employment, and disability. After that, using these keywords, a search was made to find suitable profiles and lists using (a) FHIR site search engine [16] and (b) FHIR registry [15]. The search results were as follows: (1) ValueSet EducationLevel [10]; (2) Vital Records Death Reporting (FHIR implementation guide) [21]; (3) Occupational Data for Health (FHIR implementation guide) [14]; (4) patient-disability modifier [9]; and (5) the list of disabilities [17] with SNOMED codes describing the degree of disability. The list belongs to the Quality Improvement (FHIR implementation guide) [18]. The keyword search confirmed the initial assumption that any given classifier combines three categories of data that must have been collected and processed separately. To keep the paper short and focused, we will look only at storing education and employment information.

5 RESULTS

Educational information specified as Bidirectional Services eReferral (hereinafter BSeR) [6] and Vital Records Death Reporting (hereinafter VRDR) [21]. FHIR guides implementation in the same way. It is designed as a medical record for social history in the form of FHIR Observation resource specified within BSeR_EducationLevel [7] and Decedent's Education Level [22] profiles. Both use LOINC code '80913-7 |Highest level of education [US Standard Certificate of Death]|' as an observation code, and FHIR ValueSet EducationalLevel [10] for results.

OpenEHR provides open specifications and software that can be used to build information and interoperability solutions for healthcare. Examining the openEHR specification, two archetypes can be found: (a) *Education Summary* [29], which provides summary information about an individual's current and previous education or training (Figure.1) and (b) *Education Record* [28], which specifies the individual's period of education or training.

Not all attributes in openEHR educational artefacts have mappings to terminology, and the implementer should specify these manually. The most common clinical terminology systems are SNOMED and LOINC. Examining education-related terms from both systems, the following conclusions can be drawn: (a) more than 100 different concepts are related to the SNOMED code '365458002 |Education and/or schooling finding|' [35], but these concepts do not express the level of education but rather the nature of the education or problems in receiving an education; (b) LOINC provides many education-related codes that can be used to encode openEHR

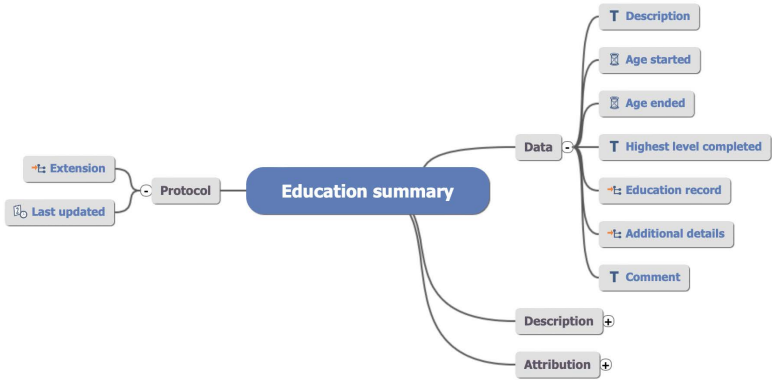

Figure 1: Mindmap of openEHR Education Summary archetype

archetypes into FHIR profiles. That being said, '82589-3 |Highest level of education|' [24] can be useful for marking the level of education. The answer options of this code can be used to express anything from preschoolers to research doctors (No schooling, 8th grade/less, High school, Doctoral degree (e.g. PhD, EdD), Unknown, and five more). Based on the structure of openEHR archetypes and LOINC coding, the Estonian profiles 'Education Level' (Figure 2) [37], 'Education Record' [39] and a list of Education Levels [38] was created, allowing both the level of education and its current status to be described.

Implementation Guide '**Occupational Data for Health**' (hereinafter ODH or ODH IG) [14] covers information about the patient's work. ODH is designed for the social history section of a medical record to facilitate clinical care in most/all disciplines and delivery environments. The FHIR observation resource is used for occupational information. The scope of the ODH information includes the following: (a) *Employment Status* – status of a person's economic relationship to work, including reference to ValueSet with values 'Employed', 'UnEmployed', 'Not In Labour Force'; (b) *Retirement Date* – a clinical statement about the date a person considers themselves 'Retired'; (c) *Combat Zone Period* – date range(s) when a person worked in what is considered a combat zone; (d) *Past or Present Job* for the patient or a household member – a clinical statement about the type of work done by an employed person; (e) *Usual Work* of the patient or a household member – a clinical statement about the type of work (paid or unpaid) done by a person for the longest time during their life, not including voluntary work.

The ODH IG covers many real-life situations, such as: (a) a patient may have their current job suspended for a combat period; (b) a patient may be retired from military service and have a current job, etc. Comparing working information in the openEHR analogously to education, two artefacts can be found: (a) 'Occupation Summary' is a summary of, or persistent information about, an individual's current and past jobs and roles [31]; (b) 'Occupation Record' is a single job or role carried out by an individual during a specified period [30].

Not all attributes in openEHR occupation artefacts have mappings to terminology, and therefore, the implementer should do terminology binding. Following that, in 'Occupation Summary', we have the 'Employment Status' attribute without a predefined ValueSet. For this purpose, the following may be selected: (a) ValueSet employmentStatusODH with values: 'Employed', 'Unemployed', 'NotInLaborForce' (persons not classified as employed or unemployed, meaning those who have no job and are not looking for one) [20]; (b) subconcepts of SNOMED concept '2365525008 |Finding of employment status|' [36], such as: 'Employed', 'In paid employment', 'Does voluntary work', 'Suspended from work', 'Not in the labour force', and more than 50 other concepts; (c) values from the answer list of LOINC concept '67875-5 |Employment status – current|' [23], such as: 'Employed full time', 'Unemployed', 'Homemaker', 'Retired due to age/preference', and four others.

We have shown that information about occupation exists both as FHIR profiles and openEHR artefacts. Employment status can be encoded using FHIR ValueSet or SNOMED/LOINC terminology. However, in our understanding, when managing information about Occupation, the FHIR ODH profiles are much more complete than the openEHR prototypes. Therefore the FHIR ODH profiles should be preferred in Estonia. The only adoption needed for ODH is referencing the national Patient, Organisation and Observation base profiles, the Estonian Register of Occupations, and terminology binding for employment status.

6 SUMMARY

As a result of this work, a migration methodology of CDA documents and their components was developed, which can be reused in the areas listed below. The methodology can be concluded in the following steps: (a) analyse each element and each list critically and deconstruct them where needed; (b) create a list of all possible search keywords and their synonyms; (c) check the FHIR site for profiles matching your search keywords; (d) check the FHIR register for profiles matching your search keywords; (e) reuse profiles, where possible; (f) search openEHR for suitable archetypes; validate their suitability and reuse, if possible; (g) search SNOMED and

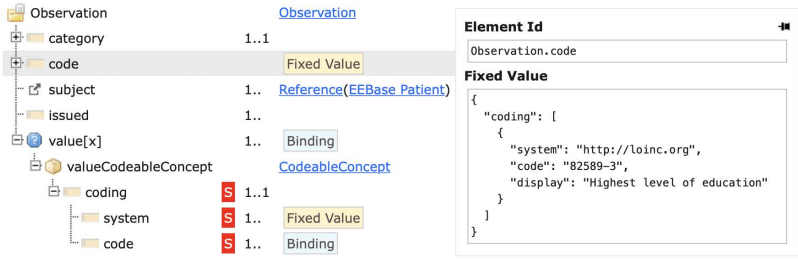

Figure 2: Estonian Education Level base profile

LOINC for matching terminology based on all search keywords; (h) develop/adapt/adopt appropriate profiles and terminology. We explained and demonstrated these steps in the paper based on real-life experiences.

Authors' contribution

I.B. designed the idea and wrote the manuscript with support from K.L. and G.P. All authors contributed to the final version. G.P. and P.R. supervised the project.

ACKNOWLEDGMENTS

This work in the project 'ICT programme' was supported by the European Union through the European Social Fund.

REFERENCES

[1] Elizabeth H Baker. 2014. Socioeconomic status, definition. *The Wiley Blackwell encyclopedia of health, illness, behavior, and society* (2014), 2210–2214.
[2] Robert H Dolin, Liora Alschuler, Calvin Beebe, Paul V Biron, Sandra Lee Boyer, Daniel Essin, Elliot Kimber, Tom Lincoln, and John E Mattison. 2001. The HL7 clinical document architecture. *Journal of the American Medical Informatics Association* 8, 6 (2001), 552–569.
[3] eHealth Network. 2017. Refined eHealth European interoperability framework. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20151123_co03_en.pdf.
[4] Lantana Group. 2022. Transitioning from CDA to FHIR. https://www.lantanagroup.com/transitioning-from-cda-to-fhir/. Accessed: 2022-07-08.
[5] Susanne Herbek, HA Eisl, Martin Hurch, Anton Schator, St Sabutsch, Günter Rauchegger, Alexander Kollmann, Th Philippi, Pia Dragon, Elisabeth Seitz, et al. 2012. The Electronic Health Record in Austria: a strong network between health care and patients. *European Surgery* 44, 3 (2012), 155–163.
[6] HL7. 2022. BSeR: Bidirectional Services eReferral. https://hl7.org/fhir/us/bser/stu1/index.html. Accessed: 2022-07-08.
[7] HL7. 2022. BSeR EducationLevel Value Set. https://hl7.org/fhir/us/bser/stu1/StructureDefinition-BSeR-EducationLevel.html. Accessed: 2022-07-08.
[8] HL7. 2022. Consolidated CDA Release. https://build.fhir.org/ig/HL7/cda-ccda-2.2/. Accessed: 2022-07-08.
[9] HL7. 2022. Extension: disability. https://www.hl7.org/fhir/R4/extension-patient-disability.html. Accessed: 2022-07-08.
[10] HL7. 2022. FHIR EducationLevel ValueSet. https://www.hl7.org/fhir/v3/EducationLevel/vs.html. Accessed: 2022-07-08.
[11] HL7. 2022. FHIR Extensibility. http://hl7.org/fhir/extensibility.html. Accessed: 2022-07-08.
[12] HL7. 2022. FHIR is a standard for health care data exchange, published by HL7®. http://hl7.org/fhir/. Accessed: 2022-07-08.
[13] HL7. 2022. FHIR Patient resource. http://hl7.org/fhir/patient.html. Accessed: 2022-07-08.
[14] HL7. 2022. FHIR Profile: Occupational Data for Health. https://hl7.org/fhir/us/odh/STU1/. Accessed: 2022-07-08.
[15] HL7. 2022. FHIR Registry. https://registry.fhir.org. Accessed: 2022-07-08.
[16] HL7. 2022. FHIR search. http://www.hl7.org/fhir/search.cfm. Accessed: 2022-07-08.
[17] HL7. 2022. QI-Core Disability Value Set. https://hl7.org/fhir/us/qicore/stu4.1.1/ValueSet-qicore-patient-disability.html. Accessed: 2022-07-08.
[18] HL7. 2022. Quality Improvement Core Implementation Guide. https://hl7.org/fhir/us/qicore/stu4.1.1/index.html. Accessed: 2022-07-08.
[19] HL7. 2022. Resource Profile: US Core Patient Profile. https://hl7.org/fhir/us/core/StructureDefinition-us-core-patient.html. Accessed: 2022-10-13.
[20] HL7. 2022. v3 Value Set employmentStatusODH. http://hl7.org/fhir/R4/v3/employmentStatusODH/vs.html. Accessed: 2022-07-08.
[21] HL7. 2022. Vital Records Death Reporting (VRDR) FHIR Implementation Guide. https://build.fhir.org/ig/HL7/vrdr/. Accessed: 2022-07-08.
[22] HL7. 2022. VRDR Decedent's Education Level. https://hl7.org/fhir/us/vrdr/StructureDefinition-VRDR-Decedent-Education-Level.html. Accessed: 2022-07-08.
[23] LOINC. 2022. Employment status - current. https://loinc.org/67875-5/. Accessed: 2022-07-08.
[24] LOINC. 2022. Highest level of education. https://loinc.org/82589-3/. Accessed: 2022-07-08.
[25] Maria Mercorella, Mario Ciampi, Massimo Esposito, Angelo Esposito, and Giuseppe De Pietro. 2016. An architectural model for extracting FHIR resources from CDA documents. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 597–603.
[26] Janek Metsallik, Peeter Ross, Dirk Draheim, and Gunnar Piho. 2018. Ten years of the e-health system in Estonia. In *CEUR Workshop Proceedings*, Adrian Rutle, Yngve Lamo, Wendy MacCaull, and Ludovico Iovino (Eds.), Vol. 2336. 3rd International Workshop on (Meta)Modelling for Healthcare Systems (MMHS), 6–15. ceur-ws.org/Vol-2336/MMHS2018_invited.pdf
[27] JLH Oei, LJGT Van Hemmen, ED Falkenberg, and S Brinkkemper. 1992. The meta model hierarchy: a framework for information systems concepts and techniques. *Department of Information Systems, University of Nijmegen* (1992).
[28] openEHR. 2022. Education Record archetype. https://ckm.openehr.org/ckm/archetypes/1013.1.3718. Accessed: 2022-07-08.
[29] openEHR. 2022. Education Summary archetype. https://ckm.openehr.org/ckm/archetypes/1013.1.3184. Accessed: 2022-07-08.
[30] openEHR. 2022. Occupation record archetype. https://ckm.openehr.org/ckm/archetypes/1013.1.3718. Accessed: 2022-07-08.
[31] openEHR. 2022. Occupation summary archetype. https://ckm.openehr.org/ckm/archetypes/1013.1.2965. Accessed: 2022-07-08.
[32] Williams Pah, Sarah Gaunt, Grahame Grieve, Vincent McCauley, and Hugh Leslie. 2012. The Development of a National Approach to CDA: Successes, Challenges and Lessons Learned in Australia. *European Journal of Biomedical Informatics* 8, 3 (2012).
[33] Christoph Rinner and Georg Duftschmid. 2016. Bridging the gap between HL7 CDA and HL7 FHIR: A JSON based mapping.. In *eHealth*. 100–106.
[34] Marten Smits, Ewout Kramer, Martijn Harthoorn, and Ronald Cornet. 2015. A comparison of two Detailed Clinical Model representations: FHIR and CDA. *European Journal of Biomedical Informatics* 11, 2 (2015).
[35] SNOMED. 2022. Education and/or schooling finding. https://browser.ihtsdotools.org/?perspective=full&conceptId1=365458002&edition=MAIN/2022-06-30&release=&languages=en. Accessed: 2022-07-08.
[36] SNOMED. 2022. Finding of employment status. https://browser.ihtsdotools.org/?perspective=full&conceptId1=365525008. Accessed: 2022-07-08.
[37] TEHIK. 2022. Education Level profile. https://simplifier.net/ee-patient/education-level. Accessed: 2022-07-08.
[38] TEHIK. 2022. Education Level Value Set. https://simplifier.net/ee-patient/eempi-education-level. Accessed: 2022-07-08.
[39] TEHIK. 2022. Education Record profile. https://simplifier.net/ee-patient/education-record. Accessed: 2022-07-08.
[40] TEHIK. 2022. Estonian Infectious Diseases Information System. https://pub.e-tervis.ee/manuals/NAKIS-TIS%20liidestusjuhend. Accessed: 2022-07-08.

Appendix 4

IV

M. Ivanova, I. Bossenko, and G. Piho, “Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution,” *Lecture Notes in Business Information Processing*, vol. 531, p. 12, 2024

Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution

Marina Ivanova[1( )], Igor Bossenko[1], Gunnar Piho[1], and Peeter Ross[2]

[1] Department of Software Science, TalTech, Akadeemia 15A, 12618 Tallinn, Estonia
marina.ivnva@gmail.com
[2] Department of Health Technologies, TalTech, Tallinn, Estonia

Abstract. In response to the evolving dynamics of healthcare, this research underscores the need for robust solutions to facilitate the exchange of clinical terminology, ensuring seamless communication and interoperability across healthcare systems. Terminology servers are pivotal in standardising and managing terminology, ensuring consistent communication and knowledge sharing. We must choose a modern, highly customisable, multilingual terminology server that supports FHIR and standard terminologies. This article offers a comprehensive overview of the challenges and limitations of existing clinical terminology exchange methods. We evaluate the strengths and weaknesses of prominent terminology servers. The findings reveal crucial insights into the current landscape of terminology management solutions, uncovering limitations and potential gaps. As a result, the article concludes with a compelling argument for the need to explore and develop a new enhanced terminology server solution. This exploration responds to the evolving demands of the modern healthcare industry and sets the stage for future advancements in clinical terminology management.

Keywords: clinical terminology exchange · terminology server · clinical terminology management · HL7 Fast Healthcare Interoperability Resources (FHIR) · Common Terminology Services 2 (CTS2)

1 Introduction

In the era of digital healthcare, the effective exchange and management of clinical information are critical for enhancing patient care, research, and overall healthcare outcomes [14]. Digitalising healthcare processes has led to diverse medical terminologies and data structures. Achieving seamless interoperability among disparate systems requires a common language for expressing clinical concepts. The success of digital medicine is contingent upon interoperability, emphasising the need for standardised clinical terminologies to bridge communication gaps between healthcare systems [18]. Clinical terminology servers play a pivotal role

M. A. Bach Tobji et al. (Eds.): ICDEc 2024, LNBIP 531, pp. 107–118, 2025.
https://doi.org/10.1007/978-3-031-76368-7_8

in this landscape by providing a standardised framework for medical vocabulary, facilitating interoperability, and ensuring accurate data representation [15].

1.1 Standards in Healthcare Data Exchange

The healthcare industry's pursuit of interoperability has led to the development of healthcare data exchange standards [18]. These standards are used every day by healthcare industry workers for record documentation, illness descriptions, etc.

HL7 Fast Healthcare Interoperability Resources (FHIR) is a modern healthcare data exchange cornerstone [5]. HL7 FHIR adopts an agile and RESTful approach, providing a flexible and efficient mechanism for sharing healthcare information over the web. Its standards-based approach promotes a common language for healthcare entities. The evolution of HL7 FHIR through its releases, including the latest R5, contributes to the adaptability and resilience of the framework, aligning it with the evolving landscape of healthcare technology [11]. The adoption of HL7 FHIR has been instrumental in addressing the challenges regarding the dependence of digital medicine on interoperability, particularly within the realm of clinical terminology [11,15].

Common Terminology Services 2 (CTS2) aims to establish a standardised interface for utilising and managing terminologies within health information technology by providing the core capabilities, functionalities, and conceptual models [10]. CTS2 connects to HL7 FHIR by providing a standardised set of models, services, and interfaces for managing and accessing terminologies [9].

1.2 HL7 FHIR Terminology Module

The HL7 FHIR (Fast Healthcare Interoperability Resources) Terminology Module is a crucial component within the HL7 FHIR standard, focused on standardising and managing the terminologies, codes, and concepts used in healthcare information exchange [12].

HL7 FHIR Terminology Module Resources:

- *CodeSystem* describes a set of codes with their meanings and relationships, specifies the rules for creating and interpreting codes within a particular domain [12].
- *ValueSet* defines a set of codes drawn from one or more code systems. Specifies the concepts or codes that can be used for a particular purpose [12].
- *ConceptMap* facilitates the mapping of codes and concepts between different code systems, enables translation of data [12].

HL7 FHIR Terminology Module Operations:

- *validate-code* checks whether a given code is valid within a specified HL7 FHIR resources [12].
- *expand* retrieves the detailed list of codes included in a ValueSet, considering any hierarchical relationships or dependencies [12].

- *lookup* provides additional information about a specific code, such as its display name, definition, and properties, and enhances the understanding of the meaning and context of a code [12].
- *translate* translates codes and concepts between different code systems using a specified ConceptMap [12].

HL7 FHIR includes a resource called "*CapabilityStatement*" that details the features and capabilities a FHIR server supports, including its interactions, search parameters, documentation, and the specific FHIR profiles and resources supported [12].

1.3 Standardised Terminology Classifications

The healthcare industry relies on a variety of standardised classifications. Standardised Terminology Classifications (STCs) form the backbone of effective communication, interoperability, and precision in exchanging clinical information. These classifications are standardised terms and codes used to represent medical concepts [3].

Standardised terminologies enhance communication, ensure data consistency, and support structured reporting within healthcare systems [3,5]. Adopting data exchange standards and standardised terminologies such as SNOMED CT [21], LOINC [19], and ICD [27] allows the ability to speak the same language in healthcare to facilitate seamless communication and data exchange.

1.4 Terminology Server

A Terminology Server (TS) in the context of medical terminology is a specialised software application designed to facilitate the creation, organisation, and exchange of standardised clinical vocabularies within the healthcare domain [16]. A TS is a centralised repository for storing and managing medical terminologies. It provides a structured environment where healthcare professionals, researchers, and system developers can access, update, and contribute to standardised clinical terms.

2 Problem

In developing a new health information system, we faced the need to use and develop terminology following the FHIR standard. The list of requirements included support for standardised terminologies, multilingual terminologies, data import/export, versioning, authoring, integration with other terminology servers and authentication systems, etc. We described the detailed criteria for comparing terminology servers in the "Terminology server criteria" section. This article addresses whether a terminology server meeting our specified requirements exists and answers the question of the strengths and weaknesses of existing terminology servers.

3 Methodology

The methodology employed for this study is rooted in a review of existing literature and resources to gather insights into the current state of clinical terminology servers and delineate the criteria for an effective server. This initial phase aimed to establish a robust foundation for the subsequent comparison of servers. We meticulously curated the requirements to encapsulate essential aspects such as multilingual support, versioning mechanisms, adherence to standards, semantic accuracy, interoperability, usability, clinical relevance, and cross-domain integration.

With this well-defined set of criteria, our study focused on acquiring a nuanced understanding of the current landscape of clinical terminology servers through a structured approach. This approach encompassed various stages, including data collection, identification of commonly used servers, brief reviews of each server, and a subsequent comprehensive comparison.

We extensively reviewed existing literature and resources to gather insights into the current state of clinical terminology servers, using search words like *clinical terminology* and *clinical terminology server*, *clinical terminology standard*, *medical terminology server*, and *FHIR terminology*. This review involved the analysis of studies, articles, and reports relevant to the utilisation and challenges associated with clinical terminology servers in contemporary healthcare settings.

A systematic and thorough comparison of the selected clinical terminology servers was subsequently conducted based on the predefined criteria. This approach ensures a comprehensive and rigorous assessment, providing valuable insights into the diverse landscape of clinical terminology servers and their alignment with the established criteria.

4 Results

In this section, we establish the criteria essential for evaluating terminology servers and delve into an overview of existing terminology server solutions. Based on the criteria list, we provide a comprehensive comparison of terminology servers.

4.1 Terminology Server Criteria

Essential criteria for evaluating a clinical terminology server should encompass various aspects to ensure effectiveness and suitability for healthcare settings. We selected the following criteria for the comparison of terminology servers:

– **Standardised terminology classification support**
A crucial criterion for evaluating clinical terminology servers is their support for standardised classification systems such as SNOMED CT and LOINC. These standards ensure consistency and interoperability in the representation of medical concepts [6,7].

- **Terminology management: CRUD operations**
  Effective terminology (CodeSystem, Concept, ValueSet, ConceptMap) management is assessed by implementing CRUD (Create, Read, Update, Delete) operations. This criterion ensures that the terminology server allows seamless manipulation of medical terms, reflecting changes and updates in clinical knowledge [15,16].
- **Terminology import**
  The ability to import terminologies from formats like CSV, TSV, FHIR JSON, FSH, ClaML XML, LOINC, and SNOMED RF2 is a practical criterion for enhancing usability and facilitating the integration of external data sources into the terminology server [1].
- **HL7 FHIR terminology module resources and operations**
  Support for HL7 FHIR terminology module resources is essential for aligning with modern interoperability standards. These resources enable the exchange of standardised healthcare information. Evaluation of terminology module operations in HL7 FHIR ensures that the terminology server aligns with the agile and RESTful approach to healthcare information exchange per HL7 FHIR standards [5].
- **Internal data model**
  The internal data model is a fundamental criterion for clinical terminology servers because it underpins the interoperability, consistency, semantic clarity, and adaptability to standards necessary for effective and efficient healthcare information exchange within a broader ecosystem. Based on widely accepted standards like CTS2 or ontology, a standardised internal data model facilitates semantic interoperability by providing a common semantic foundation for representing clinical concepts [10].
- **Versioning**
  Versioning mechanisms are critical for tracking changes in clinical terminologies over time. Versioning ensures data consistency and provides a historical perspective on the evolution of medical standards [15].
- **Multilingual terminology**
  Multilingual terminology support is essential for catering to diverse patient populations and promoting global interoperability. It ensures that terminology definitions and medical concepts can be accurately represented in multiple languages [15,16].
- **Web interface**
  The accessibility and user-friendliness of clinical terminology servers are paramount in today's digital healthcare environment. A user-friendly multilingual web interface enhances the ease with which healthcare professionals interact with and navigate the terminology system. Evaluating the web interface ensures that the clinical terminology server aligns with modern usability standards, fostering efficient utilisation by healthcare practitioners [1].
- **License type**
  The selection of a license type, whether commercial or open-source, constitutes a pivotal criterion that significantly influences the adoption and sustainability of a clinical terminology server. Open-source solutions promote

collaboration, transparency, and community-driven development, potentially leading to widespread adoption. Conversely, commercial products may offer additional features, support, and services. Evaluating the license type is crucial in understanding the terminology server's cost, flexibility, and long-term viability within the healthcare ecosystem [16].

4.2 Overview of Existing Terminology Servers

Within the dynamic landscape of healthcare informatics, diverse clinical terminology servers offer unique features and capabilities, each contributing to the ever-evolving ecosystem.

- **Ontoserver**
  Ontoserver is a syndicated terminology server designed to facilitate the efficient management and dissemination of biomedical terminologies. The primary goal of Ontoserver is to provide a centralised platform for the storage, retrieval, and distribution of biomedical terminologies, enabling standardised communication and interoperability in the healthcare and biomedical research domains. Critical features of Ontoserver include its ability to syndicate terminologies in a standardised manner, ensuring consistency and accuracy across different instances. The syndication model also supports the dynamic updating of terminologies, ensuring users can access the most current and relevant information [8,17].
- **Snowray**
  Snowray represents a web-based solution that effectively manages high-quality terminology content. Its primary purpose is to assist medical terminologists in seamlessly creating and administrating resources. Users can generate new resources or import existing ones into the platform. Within Snowray, users can effortlessly expand and maintain resources, benefiting from its user-friendly interface and useful functionalities. Besides accessing international standards, Snowray provides functionality for constructing and designing Code Systems, Value Sets, and Concept Maps from the ground up. Users can efficiently manage feedback, request changes, or engage in discussions by utilising Snowray's built-in issue system, enabling the creation of specific resource-related issues [4].
- **Rhapsody**
  Rhapsody Semantic provides a comprehensive suite of tools for achieving semantic interoperability through effective terminology management. It can seamlessly integrate into your current applications or operate independently as a standalone solution. Rhapsody features a content and subset library, auto-mapping capabilities, and a universal browsing interface, facilitating smooth interfaces and data exchange [20].
- **TermSpace**
  TermSpace is a collaborative authoring environment tailored explicitly for SNOMED CT, facilitating the seamless editing of extensions, concept creation, translation maintenance, and language localisation. The platform

boasts many features accessible through any modern browser, operating within a web-based framework and offering users an efficient, low-overhead solution. Noteworthy functionalities include the creation of new concepts in real-time and batch Quality Assurance (QA) and RF2 releases. The platform supports collaborative online efforts, allowing multiple users to work concurrently on extension maintenance [22].

- **Snowstorm**
  Snowstorm is an open-source terminology server developed by SNOMED International and designed to offer specialised support for SNOMED CT. Constructed atop Elasticsearch, its architecture prioritises optimal performance and enterprise-level scalability. Snowstorm presents two distinct APIs: the HL7 FHIR API and the Specialist SNOMED CT API. Through the HL7 FHIR API, Snowstorm facilitates integrating and utilising various code systems, including SNOMED CT, LOINC, ICD-10, and ICD-10-CM, among others. The Specialist SNOMED CT API is dedicated to managing the SNOMED CT code system, serving as a SNOMED CT Browser, and enabling the authoring of SNOMED CT editions. The emphasis on open-source architecture and Snowstorm's robust API offerings position it as a valuable asset in healthcare terminology management [13].
- **Hermes**
  Hermes encompasses a suite of terminology tools centred around SNOMED CT, featuring a high-speed terminology service equipped with robust full-text search functionality. This service is particularly well-suited for driving autocompletion in user interfaces. Additionally, Hermes incorporates an inference engine capable of analysing SNOMED CT expressions and concepts, facilitating the extraction of meaningful insights. The platform also offers cross-mapping capabilities, allowing seamless translation to and from other code systems. Hermes notably supports SNOMED CT compositional grammar and the SNOMED CT expression constraint language. Its versatile design functions as a library for integration into larger applications and as an independent, standalone microservice [26].
- **Apelon DTS**
  Apelon DTS (Distributed Terminology System) is a pivotal healthcare platform dedicated to meticulously managing and disseminating standardised clinical terminologies and associated value sets. With a primary focus on promoting interoperability and consistency, Apelon DTS offers a centralised repository for storing, retrieving, and administering diverse healthcare terminologies. This system addresses the critical need for a common framework facilitating seamless communication and data exchange across various healthcare applications. Noteworthy features include robust terminology management capabilities, versioning support for dynamic updates, and tools for mapping and cross-referencing between disparate terminologies. Apelon DTS integrates with health information systems, ensuring the uniform adoption of terminologies in electronic health records (EHRs) and other healthcare applications. The system safeguards sensitive healthcare information. Compliant with healthcare data standards, Apelon DTS upholds the use of established

terminologies such as SNOMED CT and LOINC. As an evolving solution, Apelon DTS continues to play a crucial role in enhancing the efficiency and consistency of healthcare data management [2].

4.3 Terminology Server Comparison

The Table 1 provides a structured overview of how Ontoserver, Snowray, Rhapsody, TermSpace, Snowstorm, Hermes, and Apelon DTS encompass terminology server criteria such as standardised terminology classification support, terminology management capabilities, HL7 FHIR integration, multilingual terminology support, versioning mechanisms, web interface usability, and license type.

Each server has unique features and capabilities, addressing specific aspects of clinical terminology management. The identified criteria provide a holistic view of the strengths and weaknesses of existing solutions, guiding future developments in the quest for an improved terminology server. Common strengths: standardised terminology support, CRUD operations, FHIR terminology module usage, versioning capabilities, and web interface presence. Common weaknesses: latest FHIR release support, terminology import capabilities, not standardised internal data model, and multilingual terminology support.

Table 1. Terminology server compliance with criteria

Criteria	Onto- server	Snowray	Rhapsody	Term- Space	Snow- storm	Hermes	Apelon DTS
Standardised terminology classification	SNOMED LOINC ICD	SNOMED LOINC ICD	SNOMED LOINC ICD	SNOMED	SNOMED LOINC ICD	SNOMED	SNOMED LOINC ICD
CRUD operations	+	+	+	+	N/A	+	N/A
Code System import	+	+	N/A	N/A	+	N/A	N/A
Supported FHIR release	N/A	N/A	R4	N/A	R4	N/A	N/A
FHIR terminology module resources	+	+	+	N/A	+	-	+
FHIR terminology module operations	+	+	+	N/A	+	-	+
Internal data model	N/A	N/A	N/A	N/A	own data model	N/A	N/A
Multilingual terminology	-	+	-	-	+	-	-
Versioning	+	+	+	N/A	+	+	+
Web interface	+	+	+	+	-	+	-
License type	C	C	C	C	F	F	C

Legend: + means criterion is present in the Terminology Server; - means criterion is not present in the Terminology Server; N/A means no answer available for this criteria; C means commercial license type; F means free license type.

5 Discussion

The comparative analysis of clinical terminology servers presented in this article highlights several key considerations and challenges in healthcare data exchange and interoperability. The methodology employed in evaluating these servers centred around essential criteria, such as standardised terminology classification support, terminology management, compliance with the FHIR Terminology module requirements, web user interface, multilingual terminology support, versioning mechanisms, and license type. Through this rigorous evaluation, we gained valuable insights into the strengths and weaknesses of prominent clinical

terminology servers, paving the way for a nuanced discussion on their implications for the healthcare industry.

During the evaluation process, we employed specific criteria to assess the capabilities of various servers. The examination focused on the following aspects:

- *Licence-free server*: Only servers adhering to the "free" model were considered. Snowstorm and Hermes were included in this category.
- *Rich Import Ability*: Servers demonstrating robust import ability for various file formats, including CSV, TSV, FHIR JSON, FSH, ClaML XML, LOINC, and SNOMED RF2 files, were sought. Ontoserver and Snowray were identified as matching this criterion.
- *Multilingual Web Interface*: All servers evaluated met the requirement of a multilingual web interface, except for TermSpace.
- *Web Editor and Multilingual Content Editing*: Unfortunately, no server was identified that allowed the editing of multilingual content through a web interface.
- *Multipurpose Server*: The evaluation excluded Snowstorm and TermSpace due to its limited support for terminologies, except for SNOMED CT.
- *FHIR R5 Support*: While most servers exhibited support for FHIR R4, they were only found to have support for FHIR R5. Hermes was excluded from consideration due to its lack of FHIR support during the research period.
- *Well-Known Data Model*: The assessment did not apply the criteria related to a well-known data model, as sufficient information about the data models of the servers under consideration was unavailable.

As a result, we don't find a terminology server that matches our criteria. All existing servers fail in at least two categories.

5.1 Further Work

As we navigate the intricate landscape of clinical terminology servers through the lens of our comparative analysis, certain limitations have emerged, offering distinct directions for future development. The identified challenges, including non-license-free solutions, outdated adherence to standard versions, limited capacity for importing external resources in common formats like CSV, TSV, FHIR JSON, FSH, ClaML XML, LOINC, and SNOMED RF2, compatibility with FHIR Terminology Module, and use of non-standard data models illuminate critical areas for improvement within the current ecosystem.

Exploring the feasibility of developing license-free solutions or alternative licensing models can contribute to the widespread adoption of clinical terminology servers, fostering collaboration, transparency, and community-driven development.

The imperative to align with the latest standard versions and releases represents a critical area for improvement. Future research efforts should ensure clinical terminology servers stay abreast of evolving standards, particularly in HL7 FHIR.

Enhancing the capacity for importing external resources in widely used formats such as CSV, TSV, FHIR JSON, FSH, ClaML XML, LOINC, and SNOMED RF2 is paramount. This capability ensures greater flexibility and ease of integration, allowing healthcare organisations to seamlessly incorporate external data sources into their terminology servers. Future solutions should prioritise this functionality to enhance usability and adaptability.

Aligning internal data models with the CTS2 standards is a fundamental goal for future developments. Establishing a standardised framework for terminology services, as provided by CTS2, promotes interoperability and consistency across different healthcare systems. Moving towards a unified model can bridge the current gaps and foster a more integrated approach to clinical terminology management.

In light of these challenges, a significant direction for future work involves investigating and decisively developing a new, enriched terminology server solution that addresses the identified limitations. We have decided to embark on the journey of creating a novel clinical terminology server rooted in the principles of open collaboration, adherence to the latest standards, and enhanced import capabilities. By actively addressing these areas for improvement and taking concrete steps toward creating a new solution that enriches the current landscape, we can collectively contribute to the evolution of clinical terminology management. This forward-looking approach ensures that our tools meet current needs and adapt to the evolving demands of modern healthcare, fostering a more seamless, interoperable, and standardised healthcare ecosystem.

In addition to outlining the vision and direction for developing a novel clinical terminology server, we introduce TermX - an [23] open-source solution designed to revolutionise clinical terminology management. TermX embodies our commitment to open collaboration, adherence to the latest standards, and enhanced import capabilities. For those eager to delve deeper into the intricacies of TermX and explore its functionalities, we will publish a series of articles offering comprehensive insights into its architecture and features. The source code for TermX is readily available on GitHub [24,25], providing transparency and inviting contributions from the wider community.

6 Conclusion

The comparative analysis of clinical terminology servers presented in this research underscores the pivotal role these servers play in the evolving landscape of healthcare informatics. The diverse functionalities of these clinical terminology servers cater to distinct needs within the healthcare ecosystem, offering a range of features from efficient terminology syndication to collaborative authoring environments. The comparison table is a practical resource for healthcare professionals, system developers, and researchers seeking informed guidance in selecting an appropriate clinical terminology server.

However, the analysis also revealed areas for improvement in the existing ecosystem. In the quest for an improved solution, future work should ensure

adherence to the latest standards, enhance import capabilities for flexibility, support FHIR, multilingual web viewer and editor, support multilingual content, and align internal data models with widely accepted standards. By actively addressing these areas, the healthcare industry can foster collaboration, transparency, and community-driven development, ultimately contributing to a more seamless, interoperable, standardised healthcare ecosystem.

As the healthcare landscape evolves, the insights derived from this comparative analysis inform and decisively propel our decision-making processes toward developing future clinical terminology servers. We created a new, flexible, forward-looking, open-source terminology server in response to our needs and challenges. This strategic initiative aims to lead us collectively toward a more efficient, consistent, and adaptable healthcare informatics infrastructure.

Author contributions. M.I. and I.B. designed the idea for the manuscript, and M.I. wrote it with support from I.B. All authors contributed to the final version. G.P. and P.R. supervised the project.

This work in the project "ICT programme" was supported by the European Union through the European Social Fund.

References

1. Andrews, J.C., Bogliatto, F., Lawson, H.W., Bornstein, J.: Speaking the same language: using standardized terminology (2016)
2. Apelon: Apelon Distributed Terminology System (DTS) (2024). http://www.apelondts.org/
3. Awaysheh, A., Wilcke, J., Elvinger, F., Rees, L., Fan, W., Zimmerman, K.: A review of medical terminology standards and structured reporting. J. Vet. Diagn. Invest. **30**(1), 17–25 (2018)
4. B2iHealthcare: Snowray. The Healthcare Terminology Service (2024). https://snowray.app/
5. Bender, D., Sartipi, K.: HL7 FHIR: an agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 326–331. IEEE (2013)
6. Benson, T., Grieve, G., Benson, T., Grieve, G.: Snomed CT. Principles of Health Interoperability: FHIR, HL7 and SNOMED CT, pp. 293–324 (2021)
7. Bodenreider, O., Cornet, R., Vreeman, D.J.: Recent developments in clinical terminologies-SNOMED CT, LOINC, and RxNorm. Yearb. Med. Inform. **27**(01), 129–139 (2018)
8. CSIRO: Ontoserver (2024). https://ontoserver.csiro.au/
9. Gazzarata, R., Maggi, N., Magnoni, L.D., Monteverde, M.E., Ruggiero, C., Giacomini, M.: Semantics management for a regional health information system in Italy by CTS2 and FHIR. In: Applying the FAIR Principles to Accelerate Health Research in Europe in the Post COVID-19 Era: Proceedings of the 2021 EFMI Special Topic Conference, vol. 287, p. 119. IOS Press (2021)
10. HL7: HL7 Common Terminology Services. Service Functional Model Specification. Release 2 (2022). http://www.hl7.org/documentcenter/private/standards/CTS/V3_CTS2_r2_2015FEB_R2022.pdf
11. HL7: FHIR is a standard for health care data exchange, published by HL7® (2023). http://hl7.org/fhir

12. HL7: FHIR Terminology Module (2023). http://hl7.org/fhir/terminology-module.html
13. IHTSDO: Snowstorm (2024). https://github.com/IHTSDO/snowstorm/
14. ISO: Sector: Health (2024). https://www.iso.org/sectors/health
15. Lehne, M., Sass, J., Essenwanger, A., Schepers, J., Thun, S.: Why digital medicine depends on interoperability. NPJ Digit. Med. **2**(1), 79 (2019)
16. Luna, D., Otero, C., Gambarte, M.L., Frangella, J.: Terminology services: standard terminologies to control medical vocabulary. "Words are not what they say but what they mean". In: Heston, T.F. (ed.) eHealth. IntechOpen, Rijeka (2018). https://doi.org/10.5772/intechopen.75781
17. Metke-Jimenez, A., Steel, J., Hansen, D., Lawley, M.: Ontoserver: a syndicated terminology server. J. Biomed. Semant. **9**, 1–10 (2018)
18. Health Standards Organization: Why health care standards matter (2024). https://healthstandards.org/standards/why-standards-matter
19. Regenstrief Institute, Inc.: LOINC (2023). https://loinc.org/
20. Rhapsody: Rhapsody®Semantic terminology management feature overview (2024). https://rhapsody.health/resources/rhapsody-semantic-terminology-management-feature-overview/
21. SNOMED International: SNOMED CT (2023). https://www.snomed.org/
22. Termmed: TermSpace (2024). https://cf-prod-main.termspace.com/
23. TermX: About (2024). https://termx.org
24. TermX: Download (2024). https://termx.org/download
25. TermX: Tutorial (2024). https://termx.org/docs
26. Wardle, M.: Hermes (2024). https://github.com/wardle/hermes/
27. WHO: ICD (2023). https://www.who.int/

Appendix 5

V

I. Bossenko, G. Piho, and P. Ross, “Modelling a patient identifier system in the Estonian National Health Information System,” *Lecture Notes in Business Information Processing*, vol. 531, p. 14, 2024

Modelling a Patient Identifier System in the Estonian National Health Information System

Igor Bossenko[1(✉)], Gunnar Piho[1], and Peeter Ross[2]

[1] Department of Software Science, TalTech, Akadeemia 15A, 12618 Tallinn, Estonia
igor.bossenko@taltech.ee
[2] Department of Health Technologies, TalTech, Tallinn, Estonia

Abstract. Accurate patient identification is crucial during admissions in healthcare institutions. Mistaken identity can lead to fatal consequences if patients are treated based on someone else's medical history. Identifying citizens is generally well-regulated, but accurately identifying foreign, unknown, or anonymous patients is more challenging and often lacks sufficient regulation. Our objective is to develop and assess a coding system for patient identifiers, enhancing the precision of associated health records and offering a robust, adaptable method for identifying patients from diverse backgrounds. We investigated the patient identifier system design in the Estonian National Health Information System (ENHIS) during its shift from the HL7 V3 to the HL7 Fast Healthcare Interoperability Resources (FHIR) communication protocols. This transition involved evolving from an Object Identifier (OID) system to a Uniform Resource Locator (URL) system. We devised an Identifier Domain coding system tailored for patient identification that aligns with our goals and generalised this system as a universal patient identification method. The Design Science methodology, a well-established approach in software engineering and information systems, underpins our research. We tested and illustrated our proposed patient identification coding system using examples from the Estonian Patient Register. This newly developed system enables the identification of all patient types. It is user-friendly, semantically clear, backwards compatible with the OID system, expandable, and aligns with FHIR standards. Our findings can assist in creating interoperable patient identifier systems internationally.

Keywords: Patient Identification Techniques · Patient Identity · Identifier Domain · Master Patient Index (MPI) · interoperability · Estonian National Health Information System (ENHIS) · HL7 Fast Health Interoperability Resources (HL7 FHIR) · Model-driven software engineering

1 Introduction

Utilising Unique Patient Identifiers (UPI) [38] is a prevalent practice across Europe, including in England, Wales, and the Isle of Man [14], as well as in

M. A. Bach Tobji et al. (Eds.): ICDEc 2024, LNBIP 531, pp. 132–145, 2025.
https://doi.org/10.1007/978-3-031-76368-7_10

Denmark [1], Spain [33], and Ireland [15]. This approach is also widespread outside Europe, in countries like New Zealand [37], China [3], and Israel [4]. UPI's are known for their distinctiveness, comparability, stability, and usability across various organisations [6]. Despite its advantages, the adoption of UPI's has faced policy barriers in some countries, such as the USA and Germany. In these regions, patient identification relies on documents like passports or driver's licenses [39].

Our paper aims to systematically design and explain a patient identification system that aligns with current standards and can be implemented in any national, hospital, or healthcare information system. This would support global interoperability in uniquely identifying patients.

1.1 Patient Identification in Estonian E-Health System

Our research, grounded in the documented Design Science methodology [48], examines the Estonian patient identifier system, and outlines the creation of the Estonian Patient Identifier Domain. The Estonian National Health Information System (ENHIS), a leading-edge global e-health system [9,10,30], comprehensively records health data for its citizens from birth to death.

Development of the ENHIS [35], encompassing services like discharge summaries, referrals, e-prescriptions, and electronic health records, commenced in 2005. It is built upon HL7 V3 and CDA standards, with messages defined using the HL7 Reference Information Model (RIM) and formatted in XML [41]. A pivotal element of the ENHIS is the Estonian Personal Identification Code (EstId), a national identifier for citizens and residents, essential for work, taxation, benefits, healthcare, and other government functions. The EstId is maintained by the Ministry of the Interior and issued by the Estonian Population Register.

In the ENHIS ecosystem, clinical documents reference patients using their natural identifier, such as the EstId [44].

```
<patientRole classCode="PAT">
    <id root="1.3.6.1.4.1.28284.6.2.2.1" extension="48905059995"/> ...
    <patient classCode="PSN" determinerCode="INSTANCE">
        <name> <given>Ly</given> <family>Cuusk</family> </name> ...
</patientRole>
```

Listing 1.1. Fragment of the HL7 V3 message in the ENHIS with EstId

Listing 1.1 presents a section of an HL7 V3 message within the ENHIS system, featuring an Estonian Personal Identification Code (EstId) "48905059995" [42]. This EstId format includes gender (4 for female and 20th century), year of birth (89), month (05), day (05), and a unique number (9995) for distinct identification of individuals born on the same date. This EstId appears in the "id" element's "extension" attribute. The "root" attribute signifies the namespace for EstId identifiers and is linked to an Object Identifier (OID) registry [22], adhering to the "RFC 3001 - A URN Namespace of Object Identifiers" specification [27]. The OIDs employed for patient identification are illustrated in Listing 1.2. For simplification, the ENHIS identifier (1.3.6.1.4.1.28284) is represented as "..." in this example.

```
1.3.6.1.4.1.28284 Estonian National Health Information System
 ....6 National standards
  ....6.2 Classifiers
    ....6.2.2 Technical classifiers
     ....6.2.2.1  Estonian personal identification code
     ....6.2.2.16 Personal identification codes for foreigners by country
     ....6.2.2.16.246 Personal Identification Code in Finland
     ....6.2.2.16.276 Personal Identification Code in Germany
     ...
     ....6.2.2.75 Stillborn code
    ....6.2.4 Identifiers
     ....6.2.4.7 Unknown patient identifier
```

Listing 1.2. OIDs used for patient identification in the ENHIS

For patients from abroad without an EstId, a foreign patient identification code is used, based on a document from their home country. For instance, Listing 1.3 demonstrates a foreign patient's identifier: "1.3.6.1.4.1.28284.6.2.2.16.246" is an OID for patients of the Republic of Finland, and "111111-111C" is a Finnish personal identification code [44]. An *unknown patient* refers to someone unable to provide identification at a healthcare facility, often occurring when unconscious individuals are admitted to Emergency Departments. For these patients, a unique internal identifier is assigned within the healthcare institution's namespace [44]. Listing 1.3 also shows an unknown patient's identifier, where "1.3.6.1.4.1.28284.6.2.4.7" as an OID, "90004527" representing the healthcare institution's code, and "200411111" as the internally generated patient identifier.

```
<id root="1.3.6.1.4.1.28284.6.2.2.16.426" extension="111111-123453"/>  ...(1)
<id root="1.3.6.1.4.1.28284.6.2.4.7" extension="90004527.200411111"/>  ...(2)
<id root="1.3.6.1.4.1.28284.6.2.2.75" extension="60712129993"/>        ...(3)
```

Listing 1.3. ENHIS foreign(Finnish,1), unknown (2), and stillborn (3) identifiers

1.2 Problem

Every document issued to an individual in Estonia includes an Estonian Personal Identification Code (EstId), making EstId-based patient identification in the ENHIS system highly effective for Estonian citizens. However, the Veriff database [47] indicates there are over 10,000 types of identification documents worldwide. In today's mobile world, any of these documents could serve as personal identification at Estonian healthcare institutions.

Currently, ENHIS lacks uniform rules for identifying and recording foreign patients, leading to varied practices among healthcare institutions. Some use a personal identification code, while others rely on document numbers. The diversity of documents, including various passports, ID cards, and insurance policies, further complicates the situation.

Several technical issues complicate the processing of patient IDs. One issue with the HL7 V3 and CDA documents in the ENHIS is the business restriction to use only one patient ID per document. It limits the ENHIS system from storing

multiple identifiers for the same patient. Another issue is transferring patients with unknown identifiers between healthcare institutions. Typical hospital systems generate an internal identifier for such patients (Listing 1.3), appending the healthcare institution's namespace (e.g. "90001234") only when communicating with ENHIS. However, the namespace identifying the healthcare institution is not visible to end-users and is used solely for messaging.

1.3 Research Questions

The ENHIS is currently shifting from an HL7 CDA document-based framework to an event-based model using the FHIR standard. Unlike the HL7 CDA's comprehensive patient data bundles, FHIR facilitates the immediate sharing of compact, modular resources at or right after a clinical event [46]. This transition includes establishing an independent Master Patient Index (MPI) registry to manage patient demographics, identifiers, associated individuals, and social history [43]. Our research aims to address several key questions in implementing the MPI with the FHIR standard:

1. Enhance message readability by replacing OIDs with more user-friendly identifier systems.
2. Develop a flexible and reliable system for registering foreign patients.

The paper is organised as follows: Sect. 2 details our use of the Design Science methodology to address the problems identified in Sub-sect. 1.2 and the research questions from Sub-sect. 1.3. Section 3 present the MPI solution in depth. In Sect. 4, we advocate for our proposed approach, and Sect. 5 summarises the research's contributions.

2 Method

We define several terms crucial to our discussion: A *namespace* is a context that provides a unique scope for identifiers, ensuring no clashes with identifiers outside of it [36]. An *identifier system* follows structured protocols to assign and manage unique identifiers [31]. An *object identifier* is an alphanumeric string that uniquely identifies an object within a particular system [32]. *Identification* refers to the process or techniques used to uniquely recognise entities, often using specific identifiers [38]. An *identifier domain* is a system or interconnected systems sharing a common identification method and authority [29]. Lastly, a *FHIR code system* is a collection of unique codes with associated meanings, standardised for healthcare interoperability [16].

In our research on creating a globally interoperable Master Patient Index (MPI) for the Estonian National Health Information System (ENHIS), we employed the engineering cycle from the Design Science methodology [48]. The process began with the evaluation/investigation phase, where we assessed the current patient identification challenges in ENHIS (Sub-Section 1.2) and explored

various international solutions. This was followed by the design phase, during which we developed a proposed solution based on these global practices. Subsequently, in the validation phase, we sought feedback from key Estonian stakeholders, including the Health and Welfare Information Systems Centre, the Ministry of the Interior, the Estonian Police and Border Guard Board, and the National Institute for Health Development. Having implemented the solution in ENHIS, we are now in the monitoring/evaluation phase, observing its performance in a real-life operational setting.

3 Results

Defining an identifier system for national and regional use is not difficult. In contrast, creating identifier systems for foreigners is a serious challenge since the types of identification in another country are mostly unknown (except for a passport).

3.1 Design of the Identifier Domain

Since the number of countries, document types, and healthcare institutions is large, the solution should provide a method for updating the Identifier Domain with new identifier systems when a new country, identifier type, or healthcare provider appears. The initial idea was to design the pattern of the identifier system and validate only particular segments of the identifier system. We defined three segments to support these patterns: 1) country, 2) identifier type, and 3) organisation. The ISO 3166 standard was selected for coding the countries. The FHIR IdentifierType v2-0203 [20] code system was selected for identifier types (Table 1). The country-based business or tax identifier is acceptable for the identification of organisations.

Table 2 exemplifies a multiplication of the countries and identifier types. The example contains two countries – Estonia and Germany – and two identifier types – national identifier and passport number. The identifier systems are generated automatically as a multiplication of country and identifier type. This method turned out to be too generalised. For instance, in Estonia, the passport number is not used as a patient identifier, while in Germany, there is no personal national identifier.

Therefore, we returned to the idea of a hierarchical classifier. We have deprecated the use of OID as it makes the identifier system unreadable but have kept the OID approach to creating hierarchical codes.

3.2 The URL-Based Identifier Domain

We have analysed the IANA Uniform Resource Names (URN) Namespaces registry [26] and utilised RFC3043, titled “A URN Namespace for People and Organizations” [28], to design our identifier systems. We also followed the HL7 recommendation [21] and used URLs for the identifier systems.

Table 1. List of concepts from the v2-0203 code system used in the current paper

Code	Description
NI	National unique individual identifier
PPN	Passport number
CZ	Citizenship card (ID card)
SB	Social beneficiary identifier
BCT	Birth certificate
DL	Driver's license number
PRN	Provider number
MR	Medical record number
MRT	Temporary medical record number
TAX	Tax ID number
XX	Organization identifier

Table 2. Example of generated identifier systems managed by governmental authorities

Country	Identifier type	Identifier system
Estonia (EST)	National personal identifier (NI)	EST: NI
Estonia (EST)	Passport number (PPN)	EST: PPN
Germany (DEU)	National personal identifier (NI)	DEU: NI
Germany (DEU)	Passport number (PPN)	DEU: PPN

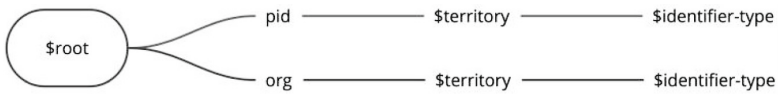

Fig. 1. Structure of the identifier system

Our identifier systems (Fig. 1) consist of (1) "$root", which is fixed in our solution as "https://fhir.ee/sid", where "https://fhir.ee" is an Estonian site for FHIR-based projects and "sid" is an abbreviation for "system of identifiers"; (2) "pid" as a namespace for patient/person identifiers or document numbers and "org" as a namespace for identification of healthcare providers; (3) the territory or country in which an identifier is issued; (4) the identifier type that allows values from the v2-0203 code system [20]. Table 3 exemplified URL-based identifier systems.

4 Discussion

The proposed methodology was adopted during development of the Estonian Master Patient Index (MPI). The Identifier Domain code system was developed first and then integrated into FHIR infrastructure of MPI.

Table 3. Examples of the URL-based patient identifiers

System URL	Description	Example of value
https://fhir.ee/sid/pid/est/ni	Estonian patient national identifier	37302102711
https://fhir.ee/sid/pid/deu/ppn	German passport number	C3JJ4789L
https://fhir.ee/sid/pid/ukr/bct	Ukrainian birth certificate	116326
https://fhir.ee/sid/pid/est/prn/90004527	Identifier issued by Parnu Hospital	123e4567-e89b a456-426614174000

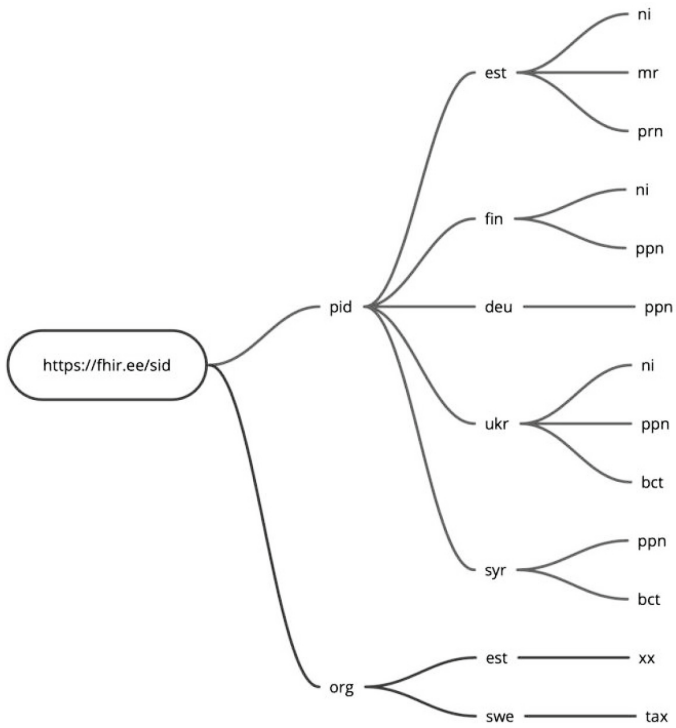

Fig. 2. Visualisation of identifier types and their subcategories in Estonia

4.1 Identifier Domain Code System

The concepts of the "Identifier Domain" code system use the developed URL notation. The visualisation of some concepts in the Identifier Domain are shown in Fig. 2. The code system has the additional properties of "identifier type", "country", "non selectable", and "naming system" [25]. Table 4 illustrates the fragment of the *Identifier Domain* code system. The "country" property refers to the issuer country from "iso3166-1-3" FHIR value set. The Identifier Domain supports patient identification based on eight identifier types: national identifier ("ni"), passport ("ppn"), ID card ("cz"), and birth certificate ("bct"), Estonian medical record number ("mr"), temporary medical number for unknown patient ("mrt"), driver licence ("dl") and provider number ("prn"). The identifier types differ by country. The "naming-system" property was added to refer to an exist-

Table 4. Fragment of the code system of the Patient Identification Domain

Lvl	Code	Display	Type	Country	Not Select-table	Naming System
1	https://fhir.ee/sid/pid/est	Estonian patient identifier namespace		EST	true	
2	https://fhir.ee/sid/pid/est/ni	Estonian personal identification code	NI	EST	false	https://fhir.ee/NamingSystem/pid-est-ni
2	https://fhir.ee/sid/pid/est/ppn	Estonian passport number	PPN	EST	false	http://terminology.hl7.org/NamingSystem/passportNumNS-EST
2	https://fhir.ee/sid/pid/est/mr	EE MPI Medical Record number	MR	EST	false	https://fhir.ee/NamingSystem/pid-est-mr
1	https://fhir.ee/sid/pid/fin	Finnish patient identifier namespace		FIN	true	
2	https://fhir.ee/sid/pid/fin/ni	Finnish personal identification code	NI	FIN	false	
2	https://fhir.ee/sid/pid/fin/ppn	Finnish passport	PPN	FIN	false	http://terminology.hl7.org/NamingSystem/passportNumNS-FIN

ing instance of the NamingSystem in Estonian or HL7 repositories. The "Not selectable" property has been added for business checks. Concepts with the "Not selectable" property equal to "false" represent an active identifier, while "true" are used only to create a hierarchy. "Not selectable" concepts like "https://fhir.ee/sid/pid/fin" are not allowed to be used to identify new patients, but may be used for search with specific-country accuracy. Listing 1.4 demonstrates the search of identifier *123456* across all Finnish identifier systems.

```
Patient?identifier = https://fhir.ee/sid/pid/fin|123456
```

Listing 1.4. Search across all Finnish patient identifiers

4.2 Identifier Domain in Master Patient Index

The *Fast Healthcare Interoperability Resource (FHIR)* facilitates healthcare information exchange, supporting multiple message formats and customisable resource definitions to suit various healthcare needs [18]. FHIR profiles can be tailored for specific contexts, including mandatory elements and terminology bindings [17]. Figure 3 exemplifies a fragment of an EEBasePatient profile [23]

Name	Flags	Card.	Type	Description & Constraints
Patient			Patient	
identifier	S	1..*	Identifier	An identifier for this patient
system	S	1..1	uri	The namespace for the identifier value **Binding:** Patient Identifier Domain (required)
value	S	1..1	string	Patient identification code or document number.
period		0..1	Period	Validity period of the identifier of identification document.
assigner		0..1	Reference(Organization)	Organization that issued the document. Can be used as a reference to an organization or as free text.

Fig. 3. FHIR Patient Identifier in the ENHIS

inherited from FHIR Patient resource [19] with mandatory attributes *identifier* and *active* (because of lower cardinality is 1). The *system* attribute is bound to *the Patient Identifier Domain* value set and restricts the use of the other values, except those defined in the bonded value set.

FHIR code systems establish specific codes and their meanings. In this study, we developed the *Identifier Domain* [25] code system. Value sets are derived from one or more code systems, defining a permissible range of codes for a given context. For instance, the *Patient Identifier Domain* value set, selected from the *Identifier Domain* code system, is tailored for patient identification. Other subsets include the *Practitioner Identifier Domain* and the *Organisation Identifier Domain*.

4.3 Previous Works

Many projects are working on patient identification, but a full solution is still elusive.

The eHealth Digital Service Infrastructure (eHDSI) is an infrastructure [8] ensuring the continuity of care for European citizens while they are travelling in the EU [7]. The eHDSI infrastructure is related to the "MyHealth@EU" brand [2], an electronic cross-border health service 25 EU countries plan to launch in 2025. The European Patient Smart Open Services (epSOS) project preceded the eHDSI between 2008 and 2014. Every eHDSI member country is a local data provider and has the legal status of data controller in that country [34]. Person health data may be queried from country "A" to country "B" ("pull" scenario) using the person identifiers supported by country "A", such as national personal identification code [5]. The health data "push" scenario from country "B" to country "A" is currently not functional [2], as in this scenario, cross-country patient identity management is one of the topics that must be solved first.

The Digital Green Certificates (DGC) for the European Union eHealth Network is a framework under which, in an interoperable manner, COVID-19 vaccination, testing and recovery certificates are issued, verified and

accepted to facilitate persons' mobility during the COVID-19 pandemic [13]. The DGC framework was designed as a part of the European eHealth Network according to Article 14 of EU Directive 2011/24/EU. The initial data set of the COVID-19 certificate contains patient identifiers [12], but was removed from the final version of the DGC framework because some EU countries could not provide it [40].

The Patient Identifier Cross-Referencing (PIX) Integration Profile is part of the technical framework of the Integrating the Healthcare Enterprise (IHE) IT Infrastructure [29]. It defines the Patient Identifier Domain as a specific healthcare information system or a space where patient identifiers are managed and associated with individual patients. The PIX profile specifies interactions and does not define any specific enterprise policies, cross-referencing algorithms, or contents the enterprise responsible for running the domain issuing patient identifiers should follow.

The European Digital Identity identifies EU citizens, residents, and businesses through electronic channels via eIDs notified under eIDAS. Using notified eIDs under the eIDAS Regulation, for the most part, will allow data providers to match an identity to an record (evidence requested) using the attributes of the natural person provided by the eIDAS minimum data set [40]. eIDs under eIDAS only cover e-channels. Furthermore, eID cannot solve the patient document-based identification problem during admission to the healthcare institution.

4.4 Changes in the Patient Identification in Estonia

The developed solution provides a flexible and reliable method for registering new identifier systems within the Identifier Domain code system based on context, country, identifier type and organisation. The critical differences in the proposed URL structure are readability due to the use of human-readable notation and the possibility of the automated issuing of identifier systems due to the regulated semantics of the identifier system. The Identifier Domain code system [24] was developed with TermX terminology server [45] and published as a part of the Estonian Base FHIR Implementation Guide.

Meetings were held with the Health and Welfare Information Systems Centre, the Ministry of the Interior, the Estonian Police and Border Guard Board, and the National Institute for Health Development to create a set of rules for developing the content of the Identifier Domain according to the country's laws. For example, the Estonian Police and Border Guard Board accepts the documents specified in the European Register of Authentic Identity and Travel Documents (PRADO) [11] for a person's identification when crossing the Estonian border. In general, PRADO includes passports from all countries over the world and ID cards from EU countries. As a result of additional research, countries were identified where all identification documents contain a national identifier. Estonia, Finland, and Latvia are examples of countries where all identification documents

contain a national identifier. For these countries, it was decided to support only the national identifier.

The new URL-based identifier system has human-readable format, designed for easy interpretation without needing extra tools. The attribute "root" in Listing 1.5(1) exemplified V3 OID-based patient identifier system, that is not human-readable. In opposite, FHIR attribute "system" (2) has notation that may be interpreted by human without additional vocabularies.

```
<id root="1.3.6.1.4.1.28284.6.2.2.1" extension="37302102711"/>          ...(1)
{"system":"https://fhir.ee/sid/pid/est/ni" value="37302102711"}         ...(2)
```

Listing 1.5. Patient identification in HL7 V3 (1) and FHIR (2) notations

The new MPI supports multiple identifiers for patients, enable linking and unlinking of patient records, including linking residents and foreigners and/or unknown patients.

5 Conclusion

This work is a unique attempt to standardise patient identification. On the one hand, the created solution enables the use of all documents identifying individuals and leverages already established identifier systems. Simultaneously, the solution outlines a methodology for describing national identifiers, emphasising their preference due to documents having a limited validity period, potentially resulting in unlinked records. Moreover, we present the content of the developed Identifier Domain for worldwide reuse. Lastly, we implemented an FHIR-based MPI and confirmed the successful integration of the generated IdentifierDomain with the FHIR integration framework.

This work may be used as a basis for the creation of the pan-European and worldwide Identifier Domain. Introducing a new global Patient Identifier Domain would facilitate a better understanding of which country or organisation and which type of identifier are used. This should improve patient identification worldwide.

Summary table

What was already known on the topic:

1. Patient identification is a global challenge with significant risks of treating someone based on someone else's medical history or lack of existing medical history.
2. There is currently no common regulation and method for identifying foreigners and unknown patients in Europe and worldwide.
3. HL7 FHIR is the most popular data exchange standard in the healthcare industry.

What this study added to our knowledge:

1. The Identifier Domain code system and Master Patient Index (MPI) as the solution for patient identification were built in Estonia.

2. Created a local regulation and method describing the addition of new identifier systems to the *identifier domain* to identify residents, foreigners, and unknown patients.
3. The human-readable URL-based notation with predefined grammar and predictable components is used in the Identifier Domain for identifier systems.
4. The developed solution is fully FHIR compatible.

Standard availability

The Identifier Domain code system has been published in the TermX terminology server [24] and as a part of the Estonian Base Implementation Guide [25].

Authors' Contribution and Acknowledgements

I.B. designed the idea and wrote the manuscript with the support of G.P. All authors contributed to the final version. G.P. and P.R. supervised the project.

This work in the project "ICT programme" was supported by the European Union through the European Social Fund.

References

1. Boyd, A., Thomas, R., Macleod, J.: NHS Number and the systems used to manage them: an overview for research users. CLOSER, London (2018)
2. Bruthans, J., Jiráková, K.: The current state and usage of european electronic cross-border health services (eHDSI). J. Med. Syst. **47**(1), 21 (2023)
3. Cheng, E.C., Le, Y., Zhou, J., Lu, Y.: Healthcare services across China-on implementing an extensible universally unique patient identifier system. Int. J. Healthc. Manag. **11**(3), 210–216 (2018)
4. Clarfield, A.M., et al.: Health and health care in Israel: an introduction. Lancet **389**(10088), 2503–2513 (2017)
5. eHealth Network: Patient Summary. Guideline on the electronic exchange of health data under Cross-Border Directive 2011/24/EU. https://health.ec.europa.eu/ehealth-digital-health-and-care/electronic-cross-border-health-services_en. Accessed 03 Feb 2024
6. Eric, H., Shan, H., Kevin, I., Andy, K., Katherine, L.: A framework for cross organizational patient identity management. The Sequoia Project and The Care Connective Consortium (2018)
7. European Commission: Patient Identifier Cross-referencing (PIX). https://health.ec.europa.eu/ehealth-digital-health-and-care/electronic-cross-border-health-services_en. Accessed 06 Oct 2023
8. European Commission: The eHealth DSI. Version 10. https://digitalhealtheurope.eu/glossary/ehdsi. Accessed 06 Oct 2023
9. European Commission: Digital Economy and Society Index (DESI) (2022). https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022. Accessed 21 Jan 2024
10. European Commission: Digital Economy and Society Index (DESI) (2023). https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts. Accessed 21 Jan 2024

11. European Council: PRADO - Public Register of Authentic identity and travel Documents Online. https://www.consilium.europa.eu/prado/en/prado-start-page.html. Accessed 15 Apr 2023
12. European Union eHealth Network: Type of Identifier for Digital Green Certificates (2021). https://health.ec.europa.eu/system/files/2021-04/digital-green-certificates_dt-specifications_en_0.pdf. Accessed 15 Oct 2023
13. European Union eHealth Network: Value Sets for Digital Green Certificates (2021). https://health.ec.europa.eu/system/files/2021-04/digital-green-certificates_dt-specifications_en_0.pdf. Accessed 15 Oct 2023
14. Fernandes, L.M., O'Connor, M.: Accurate Patient Identification: A Global Challenge. Perspectives in Health Information Management. International issue (2015)
15. Harney, C.: Implementing an individual health identifier in Ireland. Ph. D. dissertation (2012)
16. HL7: FHIR CodeSystem. hl7.org/fhir/codesystem.html. Accessed 03 Feb 2024
17. HL7: FHIR Overview - Architects. hl7.org/fhir/overview-arch.html. Accessed 15 Apr 2023
18. HL7: FHIR Overview - Developers. hl7.org/fhir/overview-dev.html. Accessed 15 Apr 2023
19. HL7: FHIR Patient. hl7.org/fhir/patient.html. Accessed 03 Feb 2024
20. HL7: HL7 IdentifierType code system. https://terminology.hl7.org/2.1.0/CodeSystem-v2-0203.html. Accessed 15 Apr 2023
21. HL7: The FHIR Identifier datatype. hl7.org/fhir/datatypes.html#Identifier. Accessed 06 May 2023
22. HL7: HL7 Object Identifier (OID) Registry (2000). hl7.org/oid/index.cfm. Accessed 15 Apr 2023
23. HL7 Estonia: EEBasePatient profile. https://fhir.ee/packages/ee-base/1.1.1/site/StructureDefinition-ee-patient.html. Accessed 03 Feb 2024
24. HL7 Estonia: The Identifier Domain at TermX terminology server. https://demo.termx.org/resources/code-systems/identifier-domain/concepts. Accessed 16 Mar 2024
25. HL7 Estonia: The Identifier Domain code system at Estonian Base Implementation Guide. https://fhir.ee/packages/terminology/0.2.2/site/CodeSystem-identifier-domain.html. Accessed 29 Jan 2024
26. IETF Network Working Group: Uniform Resource Names (URN) Syntax (1997). https://www.ietf.org/rfc/rfc2141.txt. Accessed 31 Oct 2023
27. IETF Network Working Group: A URN Namespace of Object Identifiers (2000). https://datatracker.ietf.org/doc/html/rfc3001. Accessed 15 Apr 2023
28. IETF Network Working Group: The Network Solutions Personal Internet Name (PIN): A URN Namespace for People and Organizations (2001). https://www.rfc-editor.org/rfc/rfc3043.html. Accessed 31 Oct 2023
29. IHE: Patient Identifier Cross-referencing (PIX). https://profiles.ihe.net/ITI/TF/Volume1/ch-5.html. Accessed 06 Oct 2023
30. IQVIA: Switching on the lights: Benchmarking digital health systems across emea (2022). https://www.iqvia.com/library/white-papers/switching-on-the-lights. Accessed 21 Jan 2024
31. ISO: ISO/TS 22943:2022. Information and documentation - Principles of identification. https://www.iso.org/obp/ui/en/#iso:std:iso:ts:22943:ed-1:v1:en:term:3.9. Accessed 03 Feb 2024
32. ISO: ISO/TS 28560-4:2023. Encoding of data elements based on rules from ISO/IEC 15962 in an RFID tag with partitioned memory.

iso.org/obp/ui/en/#iso:std:iso:ts:28560:-4:ed-2:v1:en:term:3.12. Accessed 03 Feb 2024
33. Kirchberger, C.: Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of crossborder eHealth. Online verfügbar unter zuletzt geprüft am **22**, 2015 (2014)
34. Linden, F.: epSOS local data providers. Acta Informatica Medica **17**(3), 142 (2009)
35. Metsallik, J., Ross, P., Draheim, D., Piho, G.: Ten years of the e-health system in Estonia. In: CEUR Workshop Proceedings, vol. 2336, pp. 6–15 (2018)
36. Oxford Reference: Namespace. https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100221995. Accessed 03 Feb 2024
37. Palliative Care Council of New Zealand: Sharing Patient Health Information (2010). https://www.moh.govt.nz/notebook/nbbooks.nsf/0/745B635139ED8647CC257E97007F46DA/$file/sharing_patient_health_information.pdf. Accessed 15 Apr 2023
38. Riplinger, L., Piera-Jiménez, J., Dooling, J.P.: Patient identification techniques-approaches, implications, and findings. Yearb. Med. Inf. **29**(01), 081–086 (2020)
39. Rosner, G.L.: Identity management policy and unlinkability: a comparative case study of the US and Germany. Ph.D. thesis, University of Nottingham (2014)
40. Schmidt, C., Krimmer, R.: How to implement the European digital single market: identifying the catalyst for digital transformation. J. Eur. Integrat. (2022)
41. Sepper, R., Ross, P., Tiik, M.: Nationwide health data management system: a novel approach for integrating biomarker measurements with comprehensive health records in large populations studies. J. Proteome Res. (2011)
42. TEHIK: Discharge summary XML message [in Estonian]. teabekeskus.tehik.ee/et/vormingud/statsionaarne-epikriis. Accessed 20 Mar 2024
43. TEHIK: Estonian Master Patient Index Implementation Guide [in estonian]. https://fhir.ee/ig/mpi/build/index.html. Accessed 03 Apr 2024
44. TEHIK: Person identifiers in ENHIS [in Estonian]. https://pub.e–tervis.ee/manuals/Juurutamisjuhend%20–%20Isiku%20koodid%20tervise%20infossteemis/1/Juurutamisjuhend-Isiku%20koodid%20tervise%20infossteemis.pdf
45. TermX: About. https://termx.org. Accessed 16 Mar 2024
46. The Office of National Coordination for Health Information Technology: What Is FHIR? https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf. Accessed 15 Apr 2023
47. Veriff: Veriff crosses the 10K line (2000). https://www.veriff.com/blog/veriff-supports-over-ten-thousand-documents. Accessed 15 Apr 2023
48. Wieringa, R.J.: Design Science Methodology for Information Systems and Software Engineering. Springer, Heidelberg (2014)

Curriculum Vitae

1. Personal data

Name	Igor Bossenko
Date and place of birth	10 February 1973, Kehra, Estonia

2. Contact information

Address	Tallinn University of Technology, School of Information Technologies, Department of Software Science, Akadeemia tee 15a, 19086 Tallinn, Estonia
E-mail	igor.bossenko@taltech.ee

3. Education

2020–2025	Tallinn University of Technology, School of Information Technologies, Computer Science, PhD
1995–1998	Tartu University, Departament of Informatics, MSc
1990–1995	Tartu University, Departament of Mathematics, BSc

4. Language competence

Russian	native
Estonian	fluent
English	fluent

5. Professional employment

2024– ...	WHO, Consultant
2020– ...	Taltech, Junior researcher
2017–2024	Kodality, CEO
2014–2017	Nortal, Chief Architect
2005–2013	Nortal, eHealth Solution Architect
2000–2005	Gennet Lab, IT Director
1994–2000	Gennet Lab, Developer

6. Field of research

- Digital Health
- Interoperability
- Terminology
- Terminology server
- Informational models

- Data transformation
- FHIR
- FHIR Mapping Language

7. Papers

1. I. Bossenko, G. Piho, and P. Ross, "Forward and Backward Compatibility Design Techniques Applying the HL7 FHIR Standard," in *HEDA@ Petri Nets*, 2022

2. R. Randmaa, I. Bossenko, T. Klementi, G. Piho, and P. Ross, "Evaluating business meta-models for semantic interoperability with FHIR resources," in *HEDA@ Petri Nets*, 2022

3. I. Bossenko, K. Linna, G. Piho, and P. Ross, "Migration from HL7 CDA to FHIR in Infectious Disease System of Estonia," in *pHealth 2022*, Studies in Health Technology and Informatics, pp. 275–278, IOS Press, 2022

4. I. Bossenko, K. Linna, G. Piho, and P. Ross, "Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pp. 882–885, 2023

5. P. Ross, J. Metsallik, K. J. I. Kankainen, I. Bossenko, C. Mäe, and M. Maasik, "Health Sense: development of a universal data model and a standard for continuity of treatment paths based on international standards of new generation health information systems," *TalTech Digikogu*, 2023. English translation: https://zenodo.org/records/14599236

6. M. Ivanova, I. Bossenko, and G. Piho, "TermX — Bridging the Gap: Implementing CTS2 and FHIR Compatible Terminology Server," in *Proceedings of Ninth International Congress on Information and Communication Technology*, vol. 1011 of *Lecture Notes in Networks and Systems*, pp. 277–287, Springer Nature Singapore, Springer, 2024

7. Bossenko, Igor and Piho, Gunnar and Ross, Peeter, "TermX: A game changer in the healthcare interoperability," in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, vol. 316 of *Studies in Health Technology and Informatics*, pp. 88–89, IOS Press, 2024

8. M. Marquis, I. Bossenko, and P. Ross, "Standardizing X-Ray and Angiography Procedure Coding in Estonia: Validating TermX Through the Implementation of SNOMED CT," in *Digital Health and Informatics Innovations for Sustainable Health Care Systems*, vol. 316 of *Studies in Health Technology and Informatics*, pp. 53–54, IOS Press, 2024

9. Bossenko, Igor and Piho, Gunnar and Ivanova, Marina and Ross, Peeter, "TermX: The semantic interoperability, knowledge management and sharing platform," *SoftwareX*, vol. 27, 2024

10. M. Ivanova, I. Bossenko, and G. Piho, "Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution," *Lecture Notes in Business Information Processing*, vol. 531, p. 12, 2024

11. I. Bossenko, G. Piho, and P. Ross, "Modelling a patient identifier system in the Estonian National Health Information System," *Lecture Notes in Business Information Processing*, vol. 531, p. 14, 2024

12. I. Bossenko, R. Randmaa, G. Piho, and P. Ross, "Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components," *Frontiers in Digital Health*, p. 30, 2024

13. M. Marquis, I. Bossenko, and P. Ross, "RadLex and SNOMED CT Integration: A Pilot Study for Standardising Radiology Classification," *Insights into Imaging*, p. 12, 2025

8. Conference presentations

1. Bossenko, Igor and Piho, Gunnar and Ross, Peeter. *Forward and backward compatibility design techniques applying the HL7 FHIR standard*. HEDA, 20-22 June 2022, Bergen, Norway.

2. Bossenko, Igor and Linna, Kerli and Piho, Gunnar and Ross, Peeter. *Migration from HL7 CDA to FHIR in Infectious Disease System of Estonia*. pHealth, 8 – 10 November 2022, Oslo, Norway.

3. Bossenko, Igor and Linna, Kerli and Piho, Gunnar and Ross, Peeter. *Migration from HL7 Clinical Document Architecture (CDA) to Fast Health Interoperability Resources (FHIR) in Infectious Disease Information System of Estonia*. Presentation and Poster. SIGAPP. 27-31 March, 2023. Tallinn, Estonia.

4. Ivanova, Marina and Bossenko, Igor and Piho, Gunnar. *Comparative analysis of clinical terminology servers: a quest for an improved solution*. ICDEc. 9-11 May, 2024. Rabat, Morocco.

5. Bossenko, Igor and Piho, Gunnar and Ross, Peeter. *Modelling a patient identifier system in the Estonian National Health Information System*. ICDEc. Presentation and session moderation. 9-11 May, 2024. Rabat, Morocco.

6. Bossenko, Igor and Piho, Gunnar and Ross, Peeter. *TermX: A Game Changer in the Healthcare Interoperability*. MIE. 24-29 August, 2024. Athena, Greece.

7. Marquis, Merit and Bossenko, Igor and Ross, Peeter. *Standardizing X-Ray and Angiography Procedure Coding in Estonia: Validating TermX Through the Implementation of SNOMED CT*. MIE. 24-29 August, 2024. Athena, Greece.

9. Presentations and workshops

1. 2023 - Organizer: Estonian Health Cluster. Location: Tallinn, Estonia. Event: Polish and Estonian cooperation. Presentation: Open-source software in Estonian National Health Information System. Role: speaker.

2. 2023 - Organizer: Estonian Health Cluster. Location: Tallinn, Estonia. Event: Tervisekogukonna Kärajad. Presentation: OpenEHR in Estonia. Role: speaker.

3. 2024 - Organizer: Technopol. Location: Cisinau, Moldova. Event: Estonian eHealth experiance. Presentation: 1) Interoperability in eHealth; 2) The architecture of Estonian National Health Information System. Role: key speaker.

4. 2024 - Organizer: Civitta. Location: Tashkent, Uzbekistan. Event: Design of Digital Health Platform. Presentation: Interoperability in eHealth. Role: speaker.

5. 2024 - Organizer: WHO. Location: Tashkent, Uzbekistan. Event: Hands-on training on the introduction HL7 FHIR to specialists from Central Asian Republics. Workshop: 5-day eHealth interoperability and FHIR training. Role: coach.

Elulookirjeldus

1. Isikuandmed

Nimi	Igor Bossenko
Sünniaeg ja -koht	10.02.1973, Kehra, Eesti
Kodakondsus	Eesti

2. Kontaktandmed

Aadress	Tallinna Tehnikaülikool (TalTech), Tarkvarateaduse Instituut, Akadeemia tee 15a, 19086 Tallinn, Eesti
E-post	igor.bossenko@taltech.ee

3. Haridus

2020–2025	Tallinna Tehnikaülikool, Infotehnoloogia teaduskond, Tarkvarateaduse instituut, PhD
1995–1998	Tartu Ülikool, Arvutiteaduse instituut, MSc
1990–1995	Tartu Ülikool, Matemaatika instituut, BSc

4. Keelteoskus

vene keel	emakeel
eesti keel	kõrgtase
inglise keel	kõrgtase

5. Teenistuskäik

2024– …	WHO, Konsultant
2020–…	Taltech, Nooremteadur
2017–2024	Kodality, Tegevjuht
2014–2017	Nortal, Peaarhitekt
2005–2013	Nortal, eHealth lahenduste arhitekt
2000–2005	Gennet Lab, IT juht
1994–2000	Gennet Lab, Arendaja

6. Teadustöö põhisuunad

- Digitaalne tervis
- Koostalitlus
- Terminoloogia
- Terminoloogiaserver
- Infomudelid
- Andmete teisendus

- FHIR
- FHIR Mapping Language

7. Teadustegevus
Teadusartiklite, konverentsiteeside ja konverentsiettekannete loetelu on toodud ingliskeelse elulookirjelduse juures.

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-258-8 (PDF)