Health Data Exchange Based on Archetypes of Clinical Concepts

Evgeniy Krastev [1], Simeon Abanos [1] and Dimitar Tcharaktchiev [2]

[1] *Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics, James Bourchier blvd., No. 5, Sofia, 1164, Bulgaria*
[2] *Medical University of Sofia, University Hospital of Endocrinology, Zdrave street No. 2, Sofia, 1431, Bulgaria*

Abstract

The electronic health record (EHR) is a core component of eHealth and the choice of the information model for representing this component is crucial for management and the quality of the healthcare services. Nowadays there are two major approaches for modeling EHR in the context of clinical data exchange. One of these approaches is represented by the HL7 set of standards. The strong sides of this approach at the level of data transmission in terms of messages over the network. However, this approach has certain weaknesses when semantic interoperability becomes a major requirement for exchange of clinical data in modern eHealth systems. This paper considers the representation of EHR in terms of ISO/EN 13606 and openEHR archetypes. The exchange of clinical data in eHealth environment using such archetypes allows achieving the highest level of interoperability among information systems in healthcare. New open platform architectures demonstrate cost efficiency in management and high quality of healthcare services using the dual information model of ISO/EN 13606 and openEHR. This paper provides results from computer experiments that demonstrate the reusability of archetypes, embedding semantic context by binding to major terminology databases, implementing constraints on data elements for ensuring quality of clinical data that is exchanged. Unlike other papers, we focus on the practical implementation of the archetypes at the production stage when instances of these archetypes become carriers of clinical data.

Keywords

eHealth, electronic health record, interoperability, clinical information models, clinical data, archetype object model, openEHR, ISO 13606

1. Introduction

Electronic healthcare (eHealth) has a growing impact on the delivery of cost-effective and secure health services through information and communication technologies. The digital processing of data improves the quality and the efficiency in healthcare, while optimizing and reducing management costs [1]. Most of the time computer processing of data in the health environment involves recording, querying and transmitting information for the purpose of decision-making about medical treatment of patients. This patient health information is represented in terms of electronic health records (EHR) and usually includes past medical history, observational data of health status and medical examinations, history, allergies as well as patient demographic details. There are many definitions of EHR, some of which are concise, while other are more detailed [2] [3]. Most of these definitions underline the need to represent, manage and share the comprehensive, structured set of health related information in EHR in accordance with widely recognized interoperability standards. This requirement becomes imperative in a globally networked world where patient health information is distributed across multiple locations, where it is usually stored in proprietary formats that are incompatible with each other [4]. Besides, a major requirement in the exchange of medical information is the preservation of the semantic context described by the individual who has authored contributions within it. Accumulating knowledge about a rapidly spreading disease like COVID-19, evaluation of drug efficiency, treatment and prevention of socially significant and rare illnesses are just few examples that illustrate the importance of semantic interoperability in EHR exchange [5] [6].

The correct interpretation of the clinical meaning of data in EHR exchange is a distinct feature of healthcare services compared to information services in other application domains. In the exiting literature, two major approaches aim to overcome this challenge. One of these approaches is represented by the HL7 set of standards for exchange of health data by means of proprietary structured messages over a computer network [7]. HL7 makes use of the application layer of the well- known Open Systems Interconnection reference model to standardize communication between different health systems. Therefore, this approach is preferred for transmission of information from sources of health information like laboratories, registries, pharmacies, finance departments to a shared repository for EHR processing. In cases when semantic interoperability is required then the ISO/EN 13606 standard [8] or the openEHR specification [9] are employed. This approach is founded on a dual reference model that separates the representation of clinical information and clinical knowledge. The syntactic and semantic capabilities of the reference model are object-oriented making its implementation suitable with modern software technologies. The reference model allows intro-

ducing archetype specifications of documents with clinical content that support terminology, security and interface considerations for the standardized exchange of EHR. In this paper, we will consider closely the development of ISO/EN 13606 and openEHR archetypes in relation to their usage in practice.

One of the obvious advantages in using archetypes is that they provide plug-and-play semantic interoperability between systems. The development of archetypes is supported by standalone or web applications such as LinkEHR, Archetype Editor or Template Designer [10] [11]. There are two major repositories for archetypes known as Clinical Knowledge Manager (CKM) [12] [13]. The CKM archetypes are built according to the openEHR specification and the CKM serves as a common platform for publishing and exchanging archetypes among the community. Each archetype is thoroughly reviewed by a team of clinicians and content experts before changing its status to "Published".

In related research work we have reported results from computer experiments with software applications of several socially significant use cases where the exchange of clinical data employs openEHR or ISO/EN 13606 archetypes of clinical documents like the International Patient Summary, the outpatient record and the medication summary of a patient [5] [14] [15] [16]. These multi-tier web applications allow web clients to manage such archetype instances on a shared EHR repository by means of RESTful web services. In this paper, we consider in detail the stage of archetype development, designing archetype templates and creating archetype instances. Moreover, we note that the quality requirements for EHR archetypes are insufficiently well studied in the context of archetype development and implementation in health information systems [17]. There are only few literature sources providing knowledge about how the ISO/EN 136060 or openEHR hierarchical reference model should be applied to represent data for a selected clinical concept. Other issues relate to selecting appropriate data types, describing events, binding terminologies and designing templates of archetypes. Another problem refers to persisting clinical data to a clinical data repository (CDR) using instances of a given archetype [18]. Unlike an archetype, the archetype instances represent concrete clinical documents containing real life data and these documents must be valid with respect to the archetype they originate. The creation of a document that is an instance of an openEHR or ISO/EN 13606 archetype template as well as the validation of a clinical document against an archetype specification are some another untrivial problems that are rather poorly explored in the literature [19] [20]. Finding a solution for these problems is essential for obtaining a software solution in terms of archetypes.

The objective of this paper is to outline the advantages in using the dual information model of ISO/EN 13606 and openEHR archetypes for describing EHR content from the viewpoint of semantic interoperability of eHealth information systems. The following section analyzes the object-oriented representa-

tion of this model and the advantages it provides in the practice for dealing with clinical concepts. New open platform architecture oriented on using openEHR specifications is discussed. Most publications on this subject focus entirely on the archetype design, while in practice the archetypes describe just the structure, the data constraints and bindings for correct interpretation of the semantic knowledge that will be exchanged once instances of these archetypes are loaded with clinical data. Here we consider the whole process of using archetypes starting with the archetype design and finish with the production stage, where the exchange and management of clinical data actually occurs. Thus, in section Results a short example is employed to illustrate in great detail the major steps that have to be followed in developing an archetype as well as the work that remains to be done after the archetype design is complete. References to software tools employed for this purpose are identified as well as the difficulties in using some of these tools. In section Conclusion we summarize the findings in this paper.

2. Methods

There are several problems in the current application domain of eHealth:
• A large amount of information is still generated on paper, which is usually then copied manually in electronic format. That further and unnecessarily prolongs the work of health professionals and at the same time increases the possibility of error.
• Clinical information systems are heterogeneous, most do not adhere to generally accepted international standards, such as ISO/EN 13606. The systems target the administration and the specific health facility, not the patient as the primary focus of care.
• The effective exchange of medical information between different systems in hospitals and clinics is difficult or missing. Moreover, if that is possible, it is usually associated with additional modifications, i.e. the information is not transferred with its semantic context. In other cases, clinical data are not available due to incompatibility of data types and structures.

One of the main challenges in medical informatics is the exchange, integration and processing of different types of information by data providers that in the general case use incompatible information models. The level of efficiency and quality of health services, as well as the management of resources in the healthcare system of each country, directly depends on finding a solution for the implementation of compatibility between these heterogeneous information systems.

There are various classifications in the literature on the level of compatibility between information systems. The lowest level of compatibility between any two such systems can be achieved by adapting the functional and syntactic

means of presenting data or the means of performing data operations. When the goal is to preserve the context of the constraints and clinical conditions that accompany the exchange of data that is achieved through semantic compatibility. In this case, when transmitting the data, their correct interpretation is guaranteed, that is their semantic correspondence from a clinical point of view.
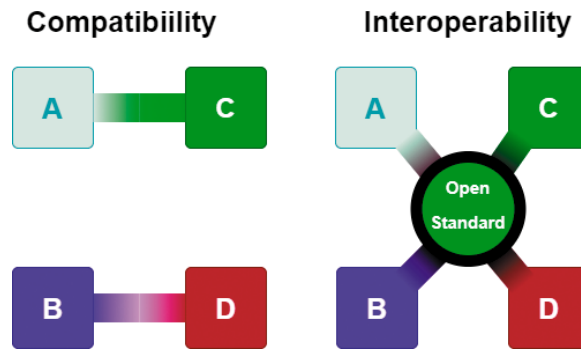

Figure 1: Difference between compatibility and interoperability

In practice, these levels of compatibility need to be implemented between any two-health information systems, and this compatibility needs to be sustainable over time, but not sporadic and inconsistent. Unlike simple compatibility between two separate systems, interoperability is based on the application of validated open standards for data presentation and clinical concepts.

There are different types of interoperability implementation classified according to the degree of automation of the process of controlling and extracting the semantic context from the exchanged data [21]:

• Functional (technical) interoperability. With this type of interoperability, data exchange is performed at the lowest level in the communication model, providing only a standard for basic data exchange services. Procedures, business processes, document flow, user cases are standardized. Characteristic of this type of interoperability is that semantic compliance cannot be controlled and implemented with digital information technologies.

• Syntactic (structural) interoperability. This type of interoperability concerns the mechanisms for packing and transmitting information. In this case, the technical compliance regarding the structure of the messages and the values contained in them is preserved, but it is not possible to guarantee or control their semantic compliance. This form of interoperability is effective where the clinical or operational objectives and the relevance of the data do not change both on the sending and receiving sides. Because the content of a structured message may not be standardized, this layer does not allow for higher levels of understanding between systems.

• Semantic interoperability. It allows for the information systems to exchange and understand the semantic context in the exchanged data. In this case, the data not only have a standard structure, but also contain coded elements to represent the semantic context in an unequivocal way. Systematized nomenclature and classification (ontologies) of diagnoses, activities and clinical condition such as SNOMED-CT [22] and ICD-11 [23] are used for coding. The coding of the semantic context allows the sharing of knowledge at the level of clinical concepts. Characteristic of this type of interoperability is that semantic compliance is perceived correctly, both by end users and through digital information technology.

Standardization of EHR is essentially required for semantic interoperability. Apparently, it is impossible to impose a restriction of the architecture design, programming language or business logic employed to build a health information system (HIS). Therefore, the only way to achieve semantic interoperability remains to introduce an information model for exchange of EHR between HIS. In fact, this is the objective of the ISO/EN 13606 standard and the openEHR specification.

Both of them use a dual information model, where a Reference model addresses information structure (Figure 2) and an Archetype Object Model defines knowledge represented in terms of archetypes. Archetypes are generic patterns describing specific properties of clinical concepts such a body mass index, body weight and height [13]. Most of the time the knowledge about clinical concept properties might change, however, the information structure remains the same. Thus, the archetypes can be updated over time without imposing any changes on the structure of information. There are minor differences in the reference models of ISO/EN 13606 and openEHR, where the reference model ISO/EN 13606 is a simplified version of the reference model openEHR. The ENTRY types of openEHR take in consideration major activities in the clinical practice. Thus, class ENTRY is a concrete class in the Reference model of ISO/EN 13606, the abstract class ENTRY in the Reference model of openEHR is specialized into concrete classes OBSERVATION, EVALUATION, INSTRUCTION and ACTION (Figure 3).

Thus, the Reference Models of ISO/EN 13606 and openEHR represent a complete object oriented design of the information entities that can be identified in the clinical practice as building blocks for describing any clinical concept in terms of archetypes [24]. The archetypes can be described in terms of XML or the Archetype Description Language (ADL). An important feature of archetypes is that they can specify constraints for data elements and bindings to terminologies such as SNOMED [22] and LOINC [25] (Figure 4). Therefore, unlike the HL7 set of standards, any instance of a clinical data can be strictly validated against the Reference Model and respectively, against the archetype it is instantiated. The

validation concerns not only the relations of the data types, the domain of data elements and, measurement units rather includes the terminology for clinical concepts, the sequence of events and the protocol used to execute a clinical activity. Respectively, the semantic context can be interpreted correctly.
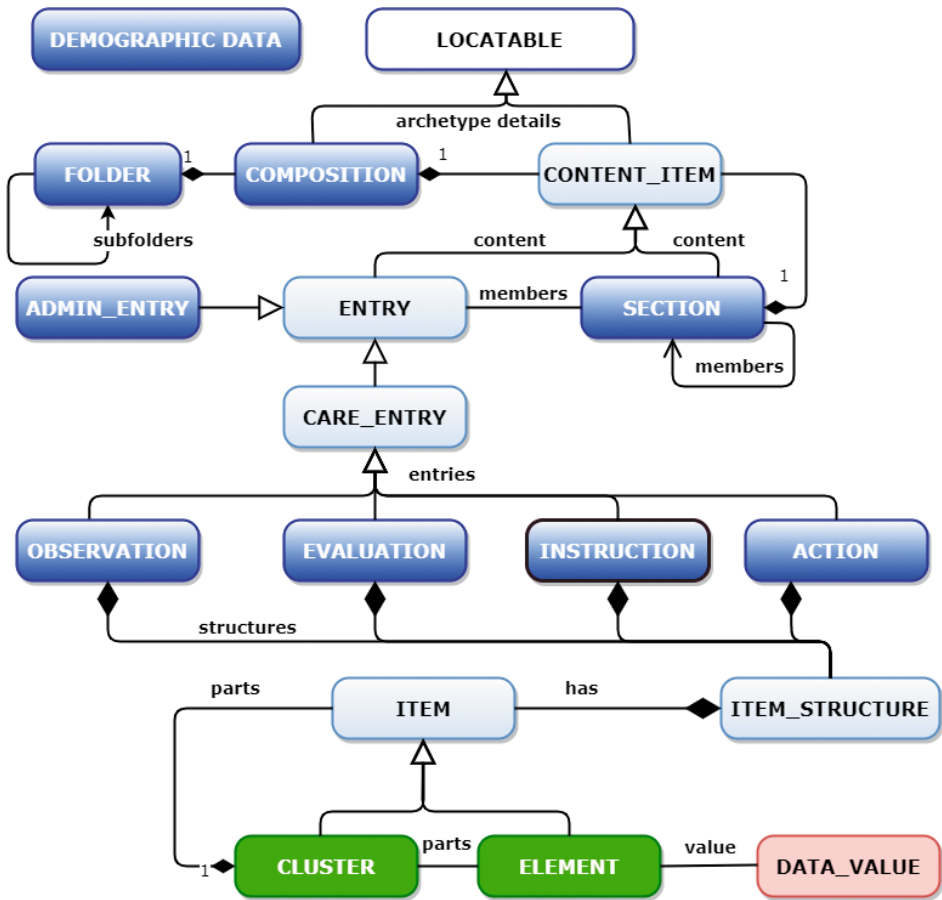

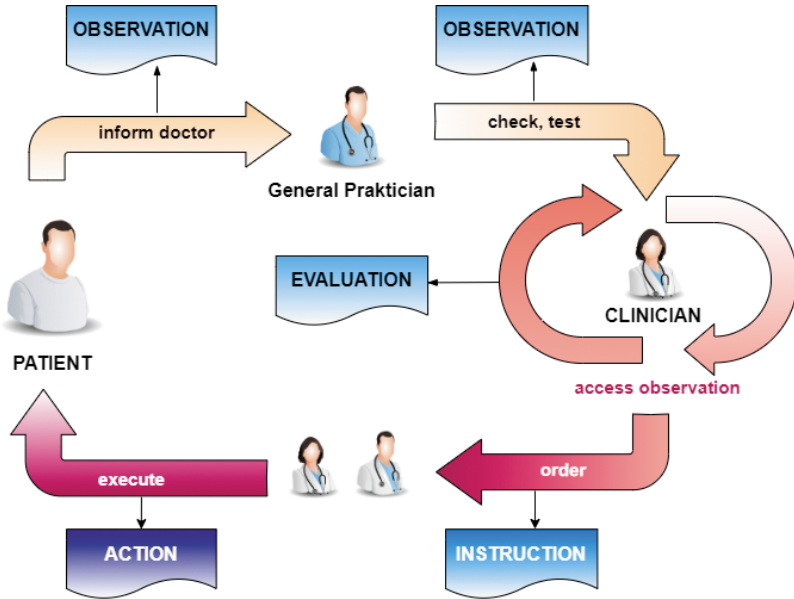**Figure 2:** UML class diagram of the Reference model of openEHR

Figure 3: Major activities in the clinical practice

Archetypes are managed by the community by means of CKM and existing archetypes can be reused as building blocks of a COMPOSITION of archetypes to represent new clinical concepts. Moreover, any COMPOSITION of archetypes builds an operational template that can be converted to an XML schema (Template Data Schema, TDS) and efficiently be used to validate XML documents with clinical data instantiated from such a XML schema using industry-standard XSD tools. Further, mappings between the archetype XML schema and potential data sources allow automatic generation of XQuery [26] scripts that translate source data into archetype compliant XML documents. The thus obtained XML documents can be managed by native XML databases like eXist-db [15] [27] or relational databases like the one use by openEHR server [18] [28]. In addition to XQuery, openEHR provides the option of using its Archetype Query Language (AQL) using archetype paths and pattern matching. Therefore, AQL are portable across physical DB schemas. A distinct feature of archetypes is their visualization in terms of mind maps that makes them easier to understand and implement in practice.
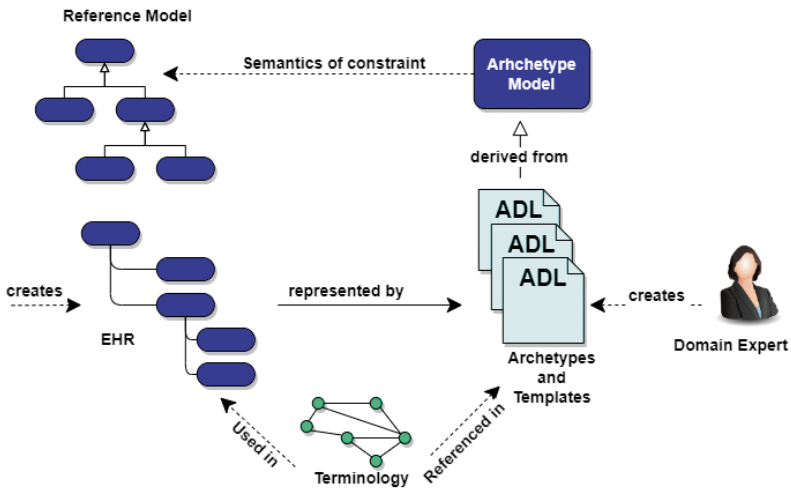
Figure 4: Relation between the Reference Model and the Archetype Object Model

One of the latest initiatives in this direction aims to develop an open platform serving as a clinical data repository (CDR), where CDR content conforms to the openEHR information model (The 5N-CDR project [29]). The CDR is managed separately from the health and social services for processing clinical data. This architecture is open for extension of such services and supports agile development of custom modules by groups or individuals of clinicians, where all modules may use multiple distributed CDRs by means of the same application-programming interface. The open platform architecture allows clinicians to define and share open-source, vendor-neutral archetypes and templates of clinical information components providing an environment to persist, process and query the data represented by these components using openEHR specifications.

3. Results

The decision about how to represent the EHR in eHealth determines the level of interoperability of information systems in the healthcare environment, the cost and an efficiency in management of this environment and finally, the quality of the healthcare services. In the section we tried to select a concise example that will demonstrates some of the major advantages in using openEHR and ISO/EN 130606 archetypes for representing EHRs, the core component of eHealth.

Let us consider the Body Mass Index (BMI) that friendly is being used in practice as an indicator for overweight and obesity. BMI is calculated using the formula and it is measured in . Measurements for and as well as the BMI calcula-

tion are done at the stage of OBSERVATION (Figure 3). Once the clinical concepts are established, we start looking for open-source archetypes that we could use. In CKM [12] we discover that the following two archetypes are published and we can reuse them:

- openEHR-EHR-OBSERVATION.body_mass_index.v2.adl (Figure 5)
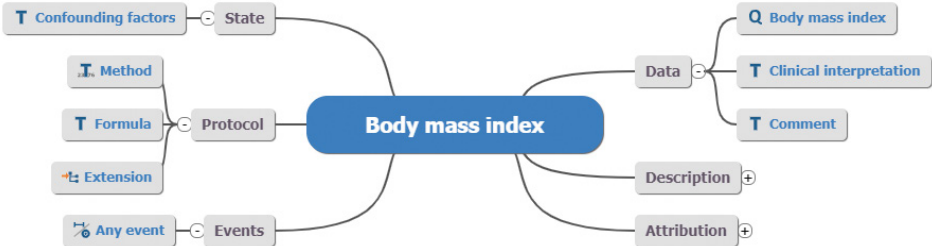- openEHR-EHR-OBSERVATION.body_weight.v2.adl (Figure 6)

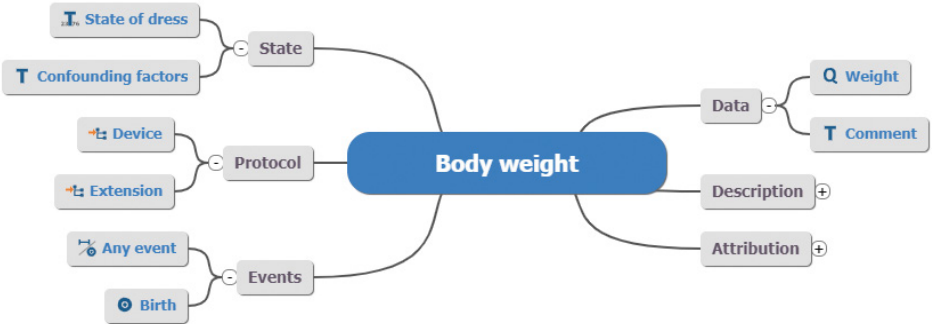**Figure 5:** Mind map of the Body Mass Index archetype

Figure 6: Mind map of the Body weight archetype

Further, we specialize the Body weight archetype into Body height archetype because an archetype for body weight is not published (approved) on CKM at the time of writing this paper (Figure 7). For this purpose, we use the Archetype Designer provided as a web application [11].
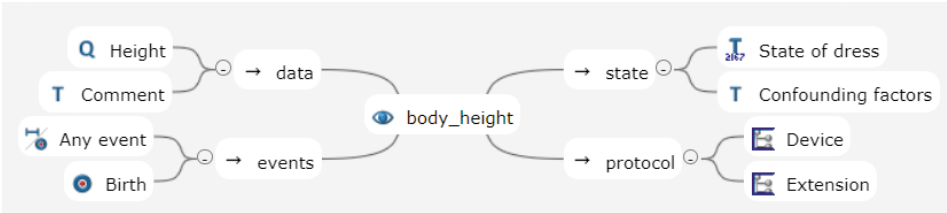

Figure 7: Mind map of the Body height archetype

Accordingly, we adjust the data properties and terminology bindings in these archetypes. The most important updates are as follows:

• The units and valid ranges of values for measurement of weight, height and BMI (Figure 8)

• The terminology server is set to SNOMED-CT and the major concepts weight, height and BMI are identified with respective SNOMED-CT codes (Figure 9)

• The Protocol and State are updated to clarify the semantic context that explains how the respective measurements are obtained. For example, weight measurement might be taken "Lightly clothed/underwear" or "Fully clothed, without shoes". On the other side, the Protocol is used to record information that may add value to the interpretation of a measurement like the device used to make the measurement. In the case of BMI, the Protocol records whether the BMI is calculated manually or by means of a digital device.

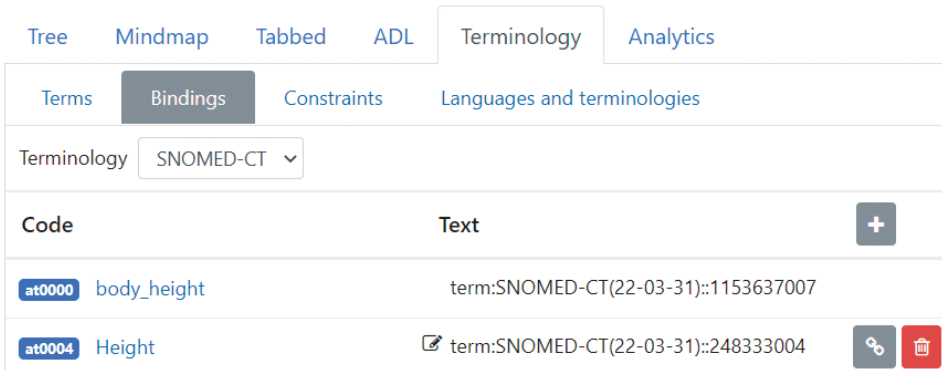**Figure 8:** Data domain validation in archetypes

Figure 9: Mind map of the Body height archetype

Finally, an archetype of type COMPOSITION is created to aggregate into a single archetype the thus obtained OBSERVATION archetypes providing Archetype slots for each one of them. This allows proceeding to the final stage of preparing the archetypes for production usage, creating an operational template that assembles the COMPOSITION archetype with the three OBSERVATION archetypes.

The production stage in using archetype involves exchanging data in electronic documents that must be valid instances of the operational template. Thus, in order to use the archetype one must load clinical data in valid instances of the operational template. Such instances might be in XML or JSON format. For this purpose, a tool is needed to export the operational template as a TDS or generates valid instances of the operational template. The development of such a tool is not trivial and for now, most of such tools are not available as open- source. In our computer experiments, we have successfully used the CaboLabs tool [19] and the Template Designer [30] to generate valid XML instances of the obtained BMI operational template, load real-life data in these instances and persist them on the openEHR server [18].

4. Conclusion

The EHR is a core component of eHealth and the choice of the information model for representing this component is crucial for management and the quality of the healthcare services. Nowadays there are two major approaches for modeling EHR in the context of clinical data exchange. One of these approaches is represented by the HL7 set of standards. The strong sides of this approach at the level of data transmission in terms of messages over the network. However, this approach has certain weaknesses when semantic interoperability becomes

a major requirement for exchange of clinical data in modern eHealth systems. This paper considers the representation of EHR in terms of ISO/EN 13606 and openEHR archetypes. The exchange of clinical data in eHealth environment using such archetypes allows achieving the highest level of interoperability among information systems in healthcare. New open platform architectures demonstrate cost efficiency in management and high quality of healthcare services using the dual information model of ISO/EN 13606 and openEHR. This paper provides results from computer experiments that demonstrate the reusability of archetypes, embedding semantic context by binding to major terminology databases, implementing constraints on data elements for ensuring quality of clinical data that is exchanged. Unlike other papers, we focus on the practical implementation of the archetypes at the production stage when instances of these archetypes become carriers of clinical data.

5. Acknowledgements

6. References

[1] G. Eysenbach, "What is e-health?," J Med Internet Res., vol. 3, no. 2:e20, 2001.

[2] The National Alliance for Health Information Technology, "Report to the Office of the National Coordinator for Health Information Technology. Defining Key Health Information Technology Terms," 28 April 2008. [Online]. Available: https://www.nachc.org/wp-content/uploads/2016/03/Key-HIT-Terms-Definitions-Final_April_2008.pdf. [Accessed 10 April 2022].

[3] Health Level Seven International, "HL7 Electronic Health Record System Functional Model,Release 2.1," 2020. [Online]. [Accessed 10 April 2022].

[4] K. Dipak, "Electronic Health Record Standards," Yearbook of medical informatics, vol. 45, no. 01, pp. 136-144, 2006.

[5] E. Krastev, D. Tcharaktchiev, K. Kaloyanova, L. Kirov, P. Kovatchev, S. Abanos and N. Mateva, "Standards Based Adaptation of Clinical Documents for Interoperability of e-Health Services", in 13[-th] conference on Information Systems and Grid Technologies (ISGT 2020), Sofia, Bulgaria, 2020, CEUR-ws.org/Vol-2656/paper2.pdf .

[6] I. Lerner, A. Serret-Larmande, B. Rance, N. Garcelon, A. Burgu, L. Chouchana and A. Neuraz, "Mining Electronic Health Records for Drugs Associated With 28-day Mortality in COVID-19: Pharmacopoeia-wide Association Study (PharmWAS)," JMIR Med Inform, vol. 10, no. 3:e35190, 2022.

[7]   Health Level 7, "HL7 Standards," Health Level 7 International, 2007. [Online]. Available: http://www.hl7.org. [Accessed 10 April 2022].

[8]   ISO 13606, "ISO 13606:2019 XML Schema. Reference model XML specifications," 2019. [Online]. [Accessed 11 April 2022].

[9]   openEHR, "EHR Information Model," The openEHR Foundation. Release 1.1.0, 29 September 2020. [Online]. Available: https://specifications.openehr.org/releases/RM/latest. [Accessed 4 April 2022].

[10] openEHR, "Modelling Tools," openEHR, 2022. [Online]. Available: https://www.openehr.org/downloads/modellingtools. [Accessed 10 April 2022].

[11] Better, "openEHR. Archetype Designer.," openEHR International, [Online]. Available: https://tools.openehr.org. [Accessed 10 April 2022].

[12] Ocean Informatics, "Clinical Knowledge Manager," 2019. [Online]. Available: https://www.openehr.org/ckm. [Accessed 10 April 2022].

[13] Apperta Foundation, "Apperta Clinical Knowledge Manager," 2020. [Online]. Available: https://ckm.apperta.org/ckm. [Accessed 10 April 2022].

[14] D. Tcharaktchiev, E. Krastev, P. Petrossians, S. Abanos, H. Kyurkchiev and P. Kovatchev, "Cross-border Exchange of Clinical Data using Archetype Concepts Compatible with the International Patient Summary", in 30th Medical Informatics Europe conference (MIE 2020), Geneva, Switzerland, pp. 552–556, 2020. Available: https://pubmed.ncbi.nlm.nih.gov/32570444.

[15] E. Krastev, D. Tcharaktchiev, P. Kovatchev and S. Abanos, "International Patient Summary Standard Based on Archetype Concepts," International Journal On Advances in Life Sciences, vol. 12, no. 1&2, p. 34:46, 2020.

[16] K. Kaloyanova, E. Krastev and E. Mitreva, "Extracting Data from General Practitioners' XML Reports in Bulgarian Healthcare to Comply with ISO/EN 13606", in 9th Balkan Conference on Informatics (BCI'19), pp. 1–5, 2019.

[17] K. Dipak, T. Archana, A. Tony and D. M. Georges, "Quality requirements for EHR Archetypes", in Quality of Life through Quality of Information, pp. 48 – 52, 2012. Available: https://ebooks.iospress.nl/publication/21702.

[18] P. P. Gutiérrez, "EHRserver," 2020. [Online]. Available: https://github.com/ppazos/cabolabs-ehrserver/releases/tag/v2.3. [Accessed 10 April 2022].

[19] G. P. Pazos, "openEHR-OPT. Java/Groovy Support of openEHR Operational Templates, Reference Model, Data Generators and other tools for www.CaboLabs.com projects," 2022.

[20] D. Boscá, J. Maldonado, D. Moner and M. and Robles, "Automatic generation of computable implementation guides from clinical information models," Journal of Biomedical Informatics, vol. 55, pp. 143–152, 2015.

[21] HIMSS, "Definition of Interoperability," 5 April 2013. [Online]. Available: https://www.himss.org/sites/hde/files/d7/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf. [Accessed 10 April 2022].

[22] SNOMED-CT, "SNOMED International," 2019. [Online]. Available: https://www.snomed.org.

[23] ICD-11, "International Classification of Diseases 11th Revision," 2018. [Online]. Available: https://icd.who.int/en.

[24] P. Muñoz, J. D. Trigo, I. Martínez, A. Muñoz, J. Escayola and J. García, "The ISO/EN 13606 Standard for the Interoperable Exchange of Electronic Health Records," Journal of Healthcare Engineering, vol. 2, no. 1, pp. 1-24, 2011.

[25] LOINC, "Logical Observation Identifiers Names and Codes," Regenstrief Institute, Inc., 2019. [Online]. Available: https://loinc.org. [Accessed 3 April 2022].

[26] W3C, "XQuery 3.0: An XML Query Language," W3C, 2019. [Online]. Available: https://www.w3.org/TR/xquery/all. [Accessed 10 April 2022].

[27] eXist Solutions, "eXistDb," 2022. [Online]. Available: http://exist-db.org. [Accessed 10 April 2022].

[28] S. Abanos, E. Krastev and D. Tcharaktchiev, "Management of Clinical Concepts in Bulgarian Healthcare Using openEHR Specifications", in The Ninth International Conference on Global Health Challenges (25–29 October), Nice, France, pp. 3–4, 2020. Available: http://www.thinkmind.org/articles/global_health_2020_1_20_78002.pdf. [Accessed October 2020].

[29] Apperta Foundation, "Defining an Open Platform," 2018. [Online]. Available: https://apperta.org/openplatforms. [Accessed 2 April 2022].

[30] Ocean Health Systems, "Template Designer," 2019. [Online]. Available: https://www.oceanhealthsystems.com/products/template-designer. [Accessed 11 April 2022].