



# The current state of summarization and visualization in Electronic Health Record (EHR) based on EHR interoperability

Amal Beldi, Salma Sassi, Richard Chbeir, Abderrazek Jemai

## ► To cite this version:

Amal Beldi, Salma Sassi, Richard Chbeir, Abderrazek Jemai. The current state of summarization and visualization in Electronic Health Record (EHR) based on EHR interoperability. Medical Information Processing and Security: Techniques and applications, Institution of Engineering and Technology, pp.87-123, 2022, 10.1049/PBHE044E\_ch5 . hal-04267193

**HAL Id: hal-04267193**

**<https://univ-pau.hal.science/hal-04267193>**

Submitted on 1 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The current state of summarization and visualization in Electronic Health Record based on EHR interoperability

Amal Beldi<sup>1</sup>, Salma Sassi<sup>2</sup>, Richard Chbeir<sup>3</sup> and Abderrazek Jemai<sup>4</sup>

<sup>1</sup>Tunis El Manar University, Faculty of Mathematical Physical and Natural Sciences of Tunis, SERCOM Laboratory, 1068 Tunis, Tunisia

<sup>2</sup>Jendouba University, Faculty of Law Economics and Management of Jendouba, VPNC Laboratory, 8189, Jendouba, Tunisia

<sup>3</sup>University Pau & Pays Adour, LIUPPA, Anglet, 64600, France

<sup>4</sup>Carthage University, Polytechnic School of Tunisia, SERCOM Laboratory, INSAT, 1080, Tunis, Tunisia

## Abstract

Recently, several healthcare organizations store heterogeneous health information about patients aiming to improve the quality of healthcare. The Electronic Health Record (EHR) contains a huge amount of patients' information making it difficult and time-saving to find the most pertinent information. Accurate, concise and automated summarization and visualization have the potential to save time, doing so includes increasing patient safety, improving efficiency, help clinical decision making, and reducing medical error as well as costs. Although interoperability and standardization are considered keys to improve the quality of care services and to coordinate care and practice effective summarization. Several studies have shown the difficulty of improving the quality of healthcare using the current summarization and visualization-based systems since they lack interoperability and they do not allow to easily express clinician needs. We found that there is no study which discusses the impact of the semantic and syntactic interoperability on EHR summarization approach which motivated us providing and discussing studies on the above topics. In this Study, we will review healthcare summarization and visualization approaches and systems, we will analyze the proposed studies according to interoperability and clinician 'needs' challenges. To construct our review, we adopted Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology and we examined papers between 1980 and 2021. Selected studies focusing on healthcare sub-areas, Electronic Health Record visualization, Electronic Health Record summarization, interoperability and standards. Based on the above papers, we provide a systematic view of development in this field and possible future directions. We conclude that most research studies in summarizing systems lack semantic interoperability and they do not rely on clinicians' needs. Besides, Electronic Health Record visualization systems lack ability to analyse efficiently health data and integrate expert knowledge domain in the decision-making process. This shall promote new research to solve these issues.

## Keywords

Electronic health record (EHR), EHR summary, summarization, semantic interoperability, standards, abstractive, extractive, visualization

---

✉ amal.beldi@fst.utm.tn (A. Beldi); salma.sassi@fsjegj.rnu.tn (S. Sassi); richard.chbeir@univ-pau.fr (R. Chbeir); abderrazekjemai@yahoo.co.uk (A. Jemai)



© 2022 Author: Please fill in the 'copyright' clause macro

CEUR Workshop Proceedings (CEUR-WS.org)

# 1. Introduction

The Information and Communication Technologies (ICT) are adopted to support healthcare services. ICT enhance the quality of care, improve healthcare service efficiencies, and enable efficient followup of care plan[1][2]. E-health is very critical mainly for patients since it facilitates and provides high-quality and safe healthcare services through accessing their relevant health information and ensures their well-being.

E-health is also important for the GP since it provides him with a decision-making system to let be informed about patient health state and eventual complications based on accessing more accurate health and accessing to medical knowledge and better practices.

It has a great interest to digitize the Electronic Health Record and all information and documents related to the patient's health. However, since Electronic health Record system contains a mix of highly heterogeneous data, the most critical component in health system still the Electronic Health Record database having different architectures that can be centralized, distributed or hybrid solutions. At the beginning, healthcare systems were isolated solutions in hospitals. After that, multiple specialized healthcare systems oriented institution or domain has appeared thus contributing itself to interoperability and information sharing issues since it seems crucial to access and to make available at anytime, anywhere all relevant clinical data and to improve the quality and delivery of care services. Interoperable Electronic Health Records have an important role for ensuring more effective and efficient patient care and facilitating the retrieval and processing of health data from heterogeneous systems. Also to provide coordinated care and practice effective communication, to aid GP decision making, and reduce medical errors and to save time. Also we interested for automated summarization and visualization challenges. Many works have discussed the topic how to summarize structured health record data in a more accessible way Thus, interoperability and standardization are considered a primordial issue to enhance and better coordinate care and practice effective summarization.

Several studies have used visualization interfaces[3];[4];[5]. Others works have interested to generate descriptions of structured time-series by using natural language . [6];[7];[8][9];[10] in order to generate a understandable summary of methods for summarizing health record data. These studies show the lack of interoperability of current summarization and visualization-based systems makes it difficult to improve the quality of care services and to easily express GP needs. Here, we conclude that most of the studies in summarizing and visualizing systems lack of semantic and syntactic interoperability and the ability to efficiently analysing health data and integrating domain expert knowledge in the decision-making process which motivated us to construct a literature review on the above topics.

This paper is organized as follow: section 2 presents research questions and motivation. Section 3 explains our reviewing methodology. In section 4, we discuss e-health syntactic and semantic interoperability issues by presenting and discussing a background of knowledge about standards, terminologies and ontologies in healthcare domain and EHR-based systems. Section 5 describes a background knowledge on EHR, reviews and discusses recent approaches on EHR summarization. In section 6 we review recent approaches on EHR visualization and discuss them into their techniques, data quality and interoperability axes. Section 7 concludes and provides some discussions on future trends.

## 2. Research Questions and Motivation

### 2.1. Motivation

Figure ?? is a summing-up diagram that demonstrates the motivation for the study of our review through visualization of keywords of the reviewed papers. Firstly, the importance of synthesis and visualization of the electronic health record to obtain brief and concise information that satisfies the user's need through the beneficial results become the primordial challenge. So we note the variety and diversity of synthesis techniques extractive, abstractive, indicative, informative and visualization tools: Natural language processing, event sequence simplification (ESS), clustering, comparison, Machine learning ML but no work has been done about the combination of summarization and visualization of the electronic health record. Except these two works [8], [11] it is interested in each process separately. Secondly, interoperability, which aims at communicating and exchanging data accurately, effectively and consistently between heterogeneous clinical systems, so through its 4 issues: Functional interoperability, technical interoperability, semantic interoperability, and syntactic interoperability remains the essential concept for enhancing the process of synthesis and the process of visualization. Thirdly The importance of various medical standards and IOT ontologies to frame and enhance a free accessibility and of electronic health record. For these reasons, we conclude to combine our study to maintain these two processes summarization and visualization for the EHR and to adopt the four interoperability techniques and challenges, to discuss different e-health standards and e-health ontologies to identify data types in the input, output process, the objective, the technique that should be used to maintain the process to orient users and the taken of different contexts from the user (patient) or the device (medical device) side.

### 2.2. Challenges

Our choice for the topic was motivated by the importance of data and systems interoperability problems to provide efficient healthcare services. E-health is a promising field to manage, predict health state and adapt care plans. So Summarization of patient information is essential practice effective communication in medicine. It 'has the potential to save time, doing so include increasing patient safety, improving efficiency, aid clinical decision making, standardize notes and reduce medical error and reducing costs. The need of making Electronic Health Record more interoperable to provide more efficient care services for patients was our first motivation to conduct this study. The purpose of this review is to provide an understanding understanding and a review of Electronic Health Record summarization and of syntactic and semantic interoperability in current Electronic Health Record systems and to provide recommendations for summarization systems based on EHR. In this paper, following challenges are addressed:

- What are the main techniques and approaches to summarize and visualize the Electronic Health Record?

- How, today, syntactic and semantic interoperability are treated in Electronic Health Record and Electronic Health Record summary?
- How to better provide syntactic and semantic interoperability in Electronic Health Records based summarization and visualization?

### 3. Reviewing methodology

In this paper we adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[3] as a reviewing methodology. To assess the quality of the selected papers, we used the Checklist for Systematic Reviews and Research Syntheses [12] of the Joanna Briggs Institute(JBI) and the qualitative research checklist [13] of Appraisal Skills Programme (CASP).

#### 3.1. Input Literature

Here, we provide details about the selection process of the chosen papers. Initially we made a two phase advance keyword search on Web of Science database. After that we conducted a search in Google Scholar and PubMed for the period 1 January 2005 to 31 December 2021 in “All Fields”. The used Keywords including “electronic Health record ” summarization ,ehr summary, data,medical data, interoperability , standard , semantic interoperability, syntactic interoperability,ontology, user profile ;were used in different phases.

#### 3.2. Methodology

To begin, we searched papers from the following conferences and journals:

- VIS: IEEE VIS conferences
- EuroVis: EuroVis conferences
- TVCG: We have carefully selected papers on EHR Vis from the IEEE Transactions on Visualization and Computer Graphics journal
- Summ: IEEE summarization conference
- VAHC: Literature published in the IEEE Workshop on Visual Analytics in Healthcare  
After analyzing the references, we found more literature that we need to conduct our review.

Fig ?? and Fig ?? summarize our methodology used for searching relevant papers. Indeed, the phase 1, thousands of papers were narrowed down. In the second phase, we found 5384 in Web of Science and PubMed. We are also conducted with the same keywords in Google Scholar search engines which resulted in 700 articles. After that, we removed duplicate articles. Finally, after eliminating duplicates, 240 articles were retained for our overview. 40 articles focus on EHR, 100 on EHR interoperability , 40 on EHR summarization and 60 on EHR visualization. Based on the outcome of the review process and on the output results we elaborated the subsequent sections.

## 4. E-health interoperability

E-health interoperability aiming at communicating and exchanging data accurately, effectively and consistently between heterogeneous clinical systems, and IoT health systems [1]. Handling heterogeneous issues is critical in EHR-based systems to ensure better communication and cooperation between health applications in order to share information through EHR and other health applications [[14][15][16]]. On the other hand there are many challenges that emerged to integrate health information systems, due to :

- i) The existence of multiple suppliers of systems and technologies,
- ii) A lot of heterogeneous clinical and administrative applications sharing information between them within the same organization, and
- iii) Each application can ensure communication through multiple and heterogeneous interfaces. Thus, achieving interoperability between clinical applications is a challenging problem since it aims at reducing health costs and contributing to more effective cares.

Ensuring the effectiveness of a clinical information systems and its interoperability with other clinical applications consists of defining and using health standards and terminologies [1]. Using interoperable systems ensuring instant access to clinical data whenever and wherever needed and avoiding clinical data redundancy in every system [2],[17] .

Several standards, terminologies and IoT based ontologies were identified to be related to either the syntactic interoperability [18] [19], or the semantic interoperability [20] or the IoT based ontologies[21],[22]. So, in next sub-sections, we firstly, provide and discuss a background of terminologies and standards. Secondly, we describe and discuss existing ontologies that can be used for the IoT domain ensuring also the interoperability between connected health systems.

### 4.1. E-health standards

Several health systems, took into account among Health Information Exchange (HIE) to ensure more appropriate, efficient, and secure access to patient's medical information. There are different classifications of interoperability that include four levels : semantic level, structural level, syntactic level and system level. ISO [23] defines two main levels: functional and semantic interoperability. In table 1 we discuss the four levels of interoperability issues classified based on the way that the information can be understood by humans and by machines.

- Functional interoperability: includes the definition of the inter-connectivity ensuring safer communicating data for one application or system.[24]
- Syntactic interoperability: includes the definition of the format of data, its syntax and how it can be organized for better interpretation[24].
- Semantic interoperability: describes a unified semantic models including normalized data and definitions to provide common understanding to the user [25]

**Table 1**

E-health Interoperability levels

Interoperability levels	Human can understand Information	Machine can understand Information
Syntactic	No	No
Functional	Yes	No
Semantic	Yes	Yes
Organizational	Yes	Yes

- Organizational interoperability: aims at simplifying a more secure and prompt communication and facilitates the use of data within and between clinical applications. [26]

We note that interoperability standards and terminologies ensure syntactic and semantic interoperability [27]. In other words, we need to standardize healthcare data to ensure interoperability and allow using it by different actors in different contexts. Healthcare applications exchange clinical information about a patient through messages. Several terminologies and standards are proposed in order to address the e-health syntactic interoperability issue such as the Health Level 7 (HL7)[28], FHIR[29], Clinical Document Architecture (CDA)[30], CEN EN 13606 EHRcom [31] and openEHR [32]. Digital Imaging and Communications in Medicine (DICOM)[33] or Statistical Data and Metadata Exchange Health Domain (SDMX-HD) [32]. These standards focus on structuring and normalizing the health data to a more accurate exchange. For the same purpose, Integrating the Healthcare Enterprise (IHE) [34] is an industry initiative specifying the Cross-Enterprise Document Sharing (XDS) to facilitate the integration profile[35]. IHE XDS aims at storing healthcare documents in an eXML registry/repository architecture and facilitating their sharing [36]. We also note the existence of many terminologies such as SNOMED-CT[37], Unified Medical Lexicon System (UMLS) [23], International Classification of Diseases (ICD)-11 [38], OpenEHR [39] and Loinc [40].

#### 4.1.1. ISO 12967:2009 (Health Informatics Service architecture)

: allows to develop interoperable e-health applications and to integrate existing ones and aims to share information within and across a health organization. This standard is composed by three parts. The first one aims to design middleware-based architecture integrating common data and business logic. [32]. The second one defines the information model and its main components to be implemented by the middleware-based architecture defined in the first level [32]. The last one aims at facilitate the exchange of information within the organization through simplifying a comprehensive and integrated interface and supporting the business processes [32].

#### 4.1.2. ISO/TS 22220:2011(Health Informatics Identification of subjects of health care)

It specifies the structure and the main data of a patient as well as the communication of his information among e-health applications. In other words, it defines demographic and other relevant data to be captured, and how they have to be implemented based on ICT-supported environment [32].



#### **4.1.3. ISO/TS 27527:2010 (Health Informatics Provider identification)**

It allows identifying and defining healthcare practitioners and healthcare organization. It intends to model the required data used to identify a general practitioner and the organization in order to correctly model their authorization and authentication in accessing health records. It also defines user roles and delegation of authority [32].

#### **4.1.4. HL7 Version 2.X.**

It is an industry standard based on a messaging exchange for application-application. It aims at covering clinical and administrative information between heterogeneous health applications. It facilitates data exchanging among heterogeneous healthcare systems. The exchanged data can be an information about admission, discharge or transfer of orders and results of laboratory tests, and clinical observations. HL7 messages are used to exchange the above data. It can also be used to transmit administrative data, ranging from appointment schedules and billing information [34]. In 1988, HL7 standard was published his first version HL7 version 1 [41]. After that, HL7 standard provided many other revisions [2]. The second one was HL7 Version 2 specifying the structure of the messages transmitted between applications. The HL7 message is a set of segments following pre-defined sequences. Segments are composed by fields which hold values for defined data types.

#### **4.1.5. HL7 Version 3**

It was developed to overcome one of the limits of HL7 Version 2 such as the lack of a flexible data model taking into account the optional existence of data elements and segments. HL7 version 3 adopts an object-oriented approach and it used the Reference Information Model (RIM) which is a static model and represents the domain of medical information. It models the grammar and semantics of HL7 Version3 messages, and uses generic classes to model health which can be easily instantiated. To develop a health applications based on HL7 v3, firstly, we define the rules used in the implementation and instantiate domain models from RIM. After that, we generate an XML diagram definitions for each message type. Thus, HL7 RIM structures envelopes supporting exchanging messages between health systems and aiming at structuring and encoding information and separating content from formatting.

#### **4.1.6. DICOM**

It defines clinical object, data structure, their semantics, and the protocols used for medical images exchanging between heterogeneous applications. It also defines the format to be used to store medical images [42].

#### **4.1.7. SDMX-HD: Statistical Data and Metadata Exchange Health Domain**

It defines the indicators, aggregates data and metadata having to be used in healthcare applications. SDMX-HD is implemented based on ISO/TS 17369:2005 SDMX standard. It aims at defining the structure and semantics of SDMX-HD messages based on XML mark-up [43].



#### **4.1.8. ASTM E2369-12: Standard Specification for Continuity of Care Record (CCR))**

It aims at structuring the summary of data including administrative information, demographic information and clinical information. Then, XML coding diagram is used to transmit summary between general practitioners and/or health services. We can visualize the constructed CCR document as a web browser, a CDA document, a PDF or word document, or in an e-mail message [34].

#### **4.1.9. HL7 Clinical Document Architecture (CDA)**

It defines a universal model of centralized EHR [44]. It also specifies the structure and semantic of clinical documents to ensure documents' interpretation by both machines and users. CDA is based on HL7 Version 3 and is the most adopted [41]. The current version of CDA, Release 2, aims at enabling human readability while being machine processable at the same time [34] by specifying the structure of the header and the body of clinical documents.

#### **4.1.10. HL7/ASTM Continuity of Care Document (CCD)**

It integrates HL7 CDA and ASTM CCR. It uses HL7 CDA to ensure reusability and interoperability by exchanging a summary of clinical data and providing a common templates for the hole parts of the summary [45].

#### **4.1.11. HL7 Care Record Summary (CRS)**

It is an application of HL7 CDA. It models the summary of various cares provided to a patient and facilitates its exchange between heterogeneous applications through a standard format [46][45].

#### **4.1.12. SNOMED-CT**

It is a medical terminology representing over 300,000 medical concepts. SNOMED-CT uses standardized and hierarchized medical concepts linked to other concepts through relationships in order to facilitate semantic interoperability [41]. It aims also at enabling the re-use of coded data for specific purposes by supporting cross mapping to other coding schemes and clinical terminology [2].

#### **4.1.13. LOINC**

It is a clinical coding system that aims at exchanging laboratory results and using a set of universal codes, names and clinical observations to facilitate the interoperability between healthcare applications [2][47].

#### **4.1.14. ICD: International Classification of Diseases**

It is an international coding system used to classify diseases, health conditions and the causes of death. It aims at coding the vital health statistics, including morbidity, mortality and medical

care reimbursement. It uses universal code disease conditions to support interoperability [48] [2].

#### **4.1.15. ICPC-2: International Classification of Primary Care, Second edition**

It is a medical classification developed by the World International Classification Committee (WICC). The WHO has adopted ICPC-2 to classify patient's diagnosis, interventions, and to model the care episode in a common structure. ICPC-2 can be used both in primary health care and general practice settings [49].

#### **4.1.16. CPT: Current Procedural Terminology**

It is a clinical terminology developed and maintained by the American Medical Association (AMA). It models medical and surgical procedures using a common coding scheme in order to simplify their exchange among healthcare institutions and general practitioners [50] [51].

#### **4.1.17. ISO 21090:2011 (Harmonized data types for information interchange)**

It enables the exchange of basic medical concepts by specifying their data types to facilitate their exchange between healthcare systems. It uses and extends the terminologies, notations and the data types defined in ISO/IEC 11404 to define and model the basic medical concepts [31].

#### **4.1.18. ISO 18308:2011 (Requirements for an electronic health record architecture)**

It models the required components for an Electronic Health Record architecture. The above architecture has to ensure the following characteristics such as reliability and clinical validation, ethical sound, and compliance with the prevailing legal requirements. EHR based architecture aims also at supporting good clinical practice, and enabling data analysis [32].

#### **4.1.19. HL7 EHR-System Functional Model, Release**

It specifies a common list of user's perspective functionalities provided in an EHR architecture in order to facilitate their reuse in a healthcare system [47].

#### **4.1.20. ISO/TS 22600 (Privilege management and access control)**

It allows the management of user privileges, ensures the access control to clinical data and supports their exchange among healthcare institutions and general practitioners [32][32][38].

#### **4.1.21. FHIR: Fast Healthcare Interoperability Resources**

It is developed to address the limits of HL7 v3 including specifically how HL7 messaging standards could be improved. FHIR provided a new approach for healthcare information exchange initially called "Resources for Health [52]. FHIR used RESTful principles [29], The FHIR is easier and more consumable than HL7 standard but it is more robust. It also uses an open Internet standards where possible.

**Table 2**

The interoperability of health standard and terminologies

	<b>Interoperability</b>			
<b>Standard</b>	<b>Technical</b>	<b>Syntactic</b>	<b>Semantic</b>	<b>Organisational</b>
<b>Identifier</b>				
ISO/TS2220:2011		*	*	
ISO/TS27527:2010		*	*	*
<b>Messaging/information exchange</b>				
HL7V2.X		*	*	
HL7V3		*	*	
DICOM		*	*	
FHIR		*	*	
SDMX HD		*		
<b>Content and structure</b>				
ASTM E2369-12		*	*	
HL7 CDA		*	*	
HL7 ASTM CCD		*	*	
HL7 CRS		*	*	
ISO21090		*	*	
<b>Health terminology</b>				
SNOMED		*	*	
LOINC		*	*	
ICD		*	*	
ICPC-2		*	*	
CPT		*	*	
electronic health record				
ISO18308:2011		*	*	*
<b>System Function Model</b>				
HL7,EHR_system functional model release1.1	*	*	*	*
<b>Security and Access Control</b>				
ISO/TS22600				*

#### 4.2. Discussion

In Table 2 we provide an overview of the purposes of the EHR standards regarding the proposed content structure. We note that the four standard are similar in terms of content structure. All

standards aiming at storing persistent and structured documents and they also support and references unstructured data specially multimedia content (images, signals and videos). Clinical data is usually stored as structured and normalized database. Each data is transformed into a standardized document to be communicated between heterogenous e-health organization, for example, in the form of HL7 messages.

### 4.3. Semantic interoperability: IoT based ontologies

The IoT health objects aims at collecting a hug amount of health data. We highlight here that this data will be multimodal structure containing multiple formats. Handling the heterogeneous data generated by various sources and processing them in real-time will be a critical in toward building connected EHR system. Several sensor ontologies described in the literature [[53],[54],[32],[55],[56],[57],[58],[43],[59],[32][2],[43],[60],[47],[48]] and aim to solve heterogeneity problems concerning the data management, the software and the hardware aspects of sensors. There are several proposed ontologies based on generic IoT concepts such as Semantic Sensor Network (SSN) ontology[53],[54],[32], M3 ontology [55] OntoSensor [34], MyOntoSens [56],[57], SAREF[58] and Brick [43]. In addition, [59] describe the places, the agents, and the events by defining a new contexts components. It provides four types of contexts classification such as external, internal, physical and logical. In order, to effectively label contextual data collected from various devices, several ontologies have been proposed for context-aware systems such as [59],[21] and [61].[33],[47], and [22] worked on the location describing the spatial context of both user and devices and estimating the location of the device through a reasoner built upon sensor data and information from location sensors. Many IoT based ontologies defined on temporal context by reusing some existing ontologies DAML-Time (DARPA Agent Markup Language project Time initiative)[2], DAML-S (DAML forWeb Services), KSL-Time (Stanford Knowledge Systems Lab Time ontology [2]) and OWL-Time ontology. To provide an universal definition of time, DAML-Time and KSL-Time defined various types of time intervals and its granularity. ontology provides concepts. Hobbs and Pan [62] proposes OWL-Time ontology describing time-based information in Gregorian calendar format. Several ontologies approaches extended and implemented OWL-Time. Another ontology was proposed by [63] aiming at annotating semantically data sensors using modular ontology Semantic Medical Sensor Data ontology (SMSD) which is built by extending multiple relevant ontologies. In this study authors propose a new model describing sensors with semantic metadata aiming at understanding their context and making them machine-understandable, interoperable, as well as facilitating data integration.

### 4.4. Discussion

To summarize, in table 3, all discussed IoT based ontologies are compared with regard to relevant aspects they manage. We propose a classification of the IoT ontologies. In order to achieve semantic consistency in different system modules the most systems propose proprietary ontologies. However, the majority of systems reuse existing ontologies to improve interoperability. Several studies combine and reuse existing ontologies to propose new ontologies for the IoT domain. IoT-Lite [64], OpenIoT ontology [65], IoT-O ontology [20], FIESTA-IoT ontology [66],

**Table 3**  
Existing IoT ontologies

Criterion	Describing Data	Discovering Sensor	Capabilities of Sensor	Accessing Data And sharing	Extensibility	Context	
						Location	Time
<b>Ontology</b>							
SSN	X		X				
Xue et al.			X	X			
Shi et al.				X			
M3	X						
OntoSensor			X				
MyOntoSens	X						
Hirmer et al.		X					
SAREF		X			X		
Brick		X			X		
Event ontology		X	X				
Baldauf et al						X	
Chen et al.						X	
Kim et al.						X	
WGS84 ontology						X	X
Flury et al.							
DAML-Time							X
KSL-Time							X
OWL-Time							X
SMSD	X		X	X	X	X	X

oneM2M ontology [67], Open-MultiNet (ONM) [13]. These new ontologies are designed for specific problems in order to collect and integrate IoT data. However, the above ontologies are still incomplete since they lack of one or more concepts for the IoT domain. Thus, we note the absence of IoT-generic ontologies defining a common ontology including core concepts. Reusing existing ontologies and merging them into new ontologies seems crucial by reusing concepts and relationships from existing ones. In order to use sensor data effectively, we identified 6 criteria aiming at choosing the appropriate ontology. The first criterion is the data description which describes the collected data to make them machine-understandable, interoperable and to facilitate data integration. Firstly, it seems important to describe the Medical Device and its capability to search specific medical contents or events. Secondly, it seems very important that the MD Network is capable to facilitate new MD implementation for time and cost savings. Thus it has to be extensible and capable to support various ways. Finally, it seems critical to propose new approaches modeling the semantics of concepts and their relationships in a particular context. For example, to correctly model the relationships between clinical concepts, we have to capture the context of medical devices by detecting the patient's activities, his location, temporal properties of the captured data and environment. As a conclusion, we provide some challenges to be addressed in order to build a semantic model taking into account the context of information and their devices:

- How to reuse existing concepts as much as possible?
- How to align them with equivalence relations?
- How to validate the constructed ontology using an ontology validator?
- How to annotate the evaluated ontology with relevant metadata?
- How to make the ontology accessible anytime and anywhere?
- How to ensure a modular approach to be sure that the ontology is reusable?
- How to and document the ontology with samples to demonstrate its usage?

The majority of the existing works in healthcare do not reuse defined IoT models. They only describe the sensor concept and ignore the description of the health data source.

## 5. Electronic Health Record Summarization

Here we describe a background knowledge on EHR, we review and discuss recent approaches on EHR summarization

### 5.1. Electronic Health Record definition

The Electronic Health Record (EHR) concept has appeared since the 1960s [68] and we note that there is no common definition of an EHR until today. Iakovidis[69] defines the EHR as a digitized clinical data about a given patient in order to support healthcare, education and research data analysis. The EHR-based application has to be accessible, secure and highly usable. In [70], Gunter and Terry define EHR as a set of clinical and electronic data about a given patient and a population. The U.S. National Cancer Institute defines EHR as an digitalized set of clinical data about a given patient stored on a computer. The U.S. Centers for Medicare and Medicaid Services defines EHR as a digitalized version of a clinical history of a given patient including the key administrative medical data and the information related to his plan cares. The EHR has to be maintained by the provider over time. The World Health Organization (WHO) [71] defines EHR as medical records provided in an EHR-based system aiming at collecting data , storing and manipulating, and providing safe access.

### 5.2. Electronic Health Record summarization approaches

The summarization process consists of creating a subset of data called summary representing the most relevant information within the original content. Existing studies on Electronic Health Record summarization focus on text summarization. In [3], authors confirm that the EHR summarization aims at providing useful information for the general practitioner by automatically create a compressed version of a given text. The type of summarized text depends on clinician's needs. However, generic summaries covering as much of the medical content included in multi-documents. The summary has to preserve the general topical organization of the original text.

In this review, we focus on the approaches based on multi-document extractive summarization which consist of producing a summary of multiple documents about the same. Summarization approaches focused on clinical variables extraction, visualization of structured and unstructured data [8], in order to provide an overview of the whole patient record. In this review, we report and discuss 38 focus papers on Electronic Health Record summarization categorized into four types of Electronic Health Record text summarization:

### 5.2.1. Extractive summaries

It consists of choosing a collection of sentences in the original document(s) to produce consistent summaries. Summaries are created by extracting phrases and sentences [72] from the original text. An extractive approach synthesizes the document(s) by identifying pieces of the patient's record and displaying the summary in user-friendly interfaces. In [73] authors perform an extractive summarization on specific diagnoses provided by the general practitioner. Their system is based on a supervised approach and uses the International Classification of Diseases (ICD) codes to train a transformer-based neural model. Radiologists made some evaluations and confirm that using supervised models generate better extractive summaries than do unsupervised approaches. In [73] authors provide a new model sighting at including accurate components to EHR data, such as structured data associated with the encounter, sentence level clinical aspects, and structures of the clinical records. Authors provide a clinical data processing pipeline based on an extension of the basic medical Natural Language Processing and on the use of concept recognition and relation detection. In [9] authors aim at customizing user views by using NLP and MedLEE NLP engine to handle modifiers. [74] defined a new method able to learn and generate meaningful topic summaries from structured clinical data. It consists of learning the correspondences between structured data and the clinical notes topics. To do so, the system uses existing summaries written by clinicians. Another study [75] provides a SIM card-based system based on medical devices. It consists of synthesizing clinical data and displaying them on the phone by a custom-developed 'Medirec' software application. [76] provided a new summarization approach aiming at summarizing text data in order to classify patients with and without diabetes. They evaluate their approach using two traditional classification methods such as logistic regression and Fisher linear discriminant analysis and four machine-learning techniques such as neural networks, support vector machines, fuzzy c-mean, and random forests. This research [77] was the unique work focusing on metastases information extraction from pathology reports of metastatic lung cancer. In [78], authors propose a new approach aiming at summarizing biomedical text documents using Bayesian summarization method. It consists of firstly mapping the input text to the Unified Medical Language System (UMLS) terminology then selecting the relevant ones to be used as classification features. [7] proposes a new approach using various graphical presentations such as small graphs for drug dosages, time lines for medical visits and hospital stays, and genograms for inherited illnesses and textual presentation of medical notes. In [8], authors propose a new approach called UPhenome. The proposed approach is based on two steps. The first one, consists of building a graphical model for large scale probabilistic phenotyping. The second step, consists of modeling diseases and patient characteristics and generate the summarized clinical data. [79] proposes a real-time summarization by aggregating clinical data from heterogeneous health care systems using HL7



messages and distributed architecture.

### 5.2.2. Abstractive summaries

Abstractive summarization techniques consist of generating new text synthesizing the original text[7]. Abstractive summaries also allow providing information context to enrich the data. In [80],[81] a first method called Timeline consists of involving clinicians to code rules while AdaptEHR consists of inferring rules automatically and relationships from ontologies and graphical models. [82] proposes a hybride method based on abstractive and extractive summurization of clinical variables. It aims at performing semantic, temporal, and contextual abstraction using domain-specific ontology to generate abstractions. Finally, [83],[13] proposed a new method to graphically summarize clinical data by generating new text.

### 5.2.3. Indicative summaries

Indicative Summurization approach extracts significant term of the original text and highlights the main parts. Indicative summaries are used in conjunction with EHR, in order to integrate indirectly in the extractive summarization process. There are few approaches in the literature concerning indicative summarization. [72] proposed a new approach to summurize and graphically visualize the EHR. [72] proposed a task-based evaluation summarizer. In [84] authors evaluated how and when clinicians in an ambulatory setting will enter data directly into an EHR.

### 5.2.4. Informative summaries

Informative summurization approach designs summaries to be used independently of the EHR. The informative approach is used as to replace the original set of raw data [85],[80]. [85] proposes a new method summarizing structured clinical data. such as administrative, computerized provider order entry and laboratory test data. Authors developed a new model to detect risks by predicting each of 2 severity levels of in-hospital AKI. Other studies using the visualization-based summarization approach [86];[87] ;[88] and RDF based summarization approach [89] are proposed to summarize text data. Many research groups[72] (NUCRSS) ;[10] [90], proposed clinical data summarizations system based on the text input data. The proposed system aims to reduce data volume. Also, many frameworks are proposed for text summarization [91];[92];[93];[52] and for generating new stories and scientific articles to summarize unstructured texts [94],[95], [91],[96]. In order to overcome challenges relating to EHR summarization and to compare and discuss existing studies in clinical data summarization, all criteria defined in the previous section were used to characterize and compare clinical data summarization methods according to the challenges highlighted in the introduction section. Our analysis summarized in Table 5 demonstrates that the most of clinical summarization approaches are based on unstructured data [73],[97] ;[77];[98] and rely on structure [76] ;[99], [100],[101] and text content in order to construct the summary. Also, our comparison highlights that the evolution of the summary is still an open challenge. So, we observe that none of the existing studies consider real data in their analysis and do not consider the context on creating the summary and few of them [100],[101],[102] rely only on the time property. Thus, existing

systems are still unable to contextually interpret and reason on the transferred knowledge among real data, and consequently cannot synthesize data to provide accurate desired results. All existing systems focus on one or two objectives at most, while none of them provide in the same framework various functionalities despite its importance in supporting Clinicians' preferences to find the patients data according to various needs. All objectives should be an integral part of a clinical summarization-based system. Finally, another important point is the output type of summarized clinical data. The output of most of the existing systems is textual format. Two of them propose a graphical summarization [100];[75]. They neither propose dedicated tools that make the summary easier to be understood and interpreted by the clinician nor provide them with appropriate perceptions of their needs. An intuitive and friendly GUI will benefit the data summarization-based systems.

Table 4: Summary of retrieved studies on electronic health record and medical data summarization

Approach	Input data	Data type	Standard	Output	Context	User Oriented	Summarization Technique	Summarization Approach
Denis Jered et al 2020	Unstrufted	Text	No	Text	No	No	Classification	Extractive
Liang et al 2019	Unstructured	Document	No	Document	No	No	Classification	Extractive
P Durgaetal 2018	Unstructured	Numerical	No	Document	Time	No	Classification	Extractive
Jen et al 2018	Structured	Text	No	Text	Time	No	Classification Filtering	Extractive
Soysal et al. 2017	Unstructured	Text	No	Text	No	No	Visualization	Extractive
Moradi and Ghadiri 2017	Unstructured	Text	Yes	Document	No	No	Extraction	Extractive
Razavian et al. 2015	Unstructured	Text	No	Text	No	No	Analysis	Extractive
Borland 2014	Structured	Text	No	Graph	Time	No	Visualization	Extractive
Fei et al. 2013	Structured	Numerical	No	Document	Time	No	Analysis	Extractive
Antonelli et al 2013	Structured	Numerical	No	Document	No	No	Visualization	Extractive
Tapak et al. 2013	Structured	Numerical	No	Document	No	No	Analysis Classification	Extractive
Klann et al. 2013	Unstructured	Document	No	Document	No	No	Analysis	Extractive
Roque et al. 2010	Unstructured	Text	No	Text	No	No	Extraction	Extractive
Rind et al. 2010	Unstructerd	Text	No	Text	No	No	Analysis Visualization	Extraction
Krummenacher et al. 2010	Structured	XML	Yes	Tuple	No	No	Visualization	Extraction
were.C et al. 2010	Unstructured	Text	No	Text	Time	No	Extraction Indicative	Extraction Abstraction
Barakat et al. 2010	Unstructured	Text	No	Graphic	No	No	Abstraction Informative	Abstraction
Savova et al. 2010v	Unstructured	text	No	Text	No	No	Analysis	extraction
Krummenacher et al. 2009	Unstructured	XML	Yes	Document	No	No	Visualisation	Extractive Abstractive
Kumar and al. 2008	Unstructured	Text	No	Text	No	No	Extraction	Extractive

Huang et al. 2007	Unstructured	Text	No	Text	No	No	Clustering	Extractive
Reeve et al. 2007	Unstructured	Text	Yes	Text	No	No	Extraction	Extractive
Afantenos 2006	Unstructured	Text	Yes	Text	No	No	Extraction	Extractive
Ogers et al. 2006	Unstructured Structured	Text	Text	No	No	No	Aggregation	Extractive
Reeve et al. 2006	Unstructured	Text	Yes	Text	No	No	Extraction	Extractive
Bui et al. 2007	Structured Unstructured	Xml	Yes	Text Image	Time	No	Extraction Abstraction	Extractive Abstractive
Wilcox et al. 2005	Unstructured	Text	No	Text	No	No	Extraction	Extractive
Liu and Friedman 2004	Structured	Xml	No	Text	No	No	Extraction	Extractive
Payan 2004	Unstructured	Text	No	Text	No	No	Extraction	Abstractive
Radev'hovy and McKeown 2002	Unstructured	Text	No	Text	No	No	Extraction	Abstractive
Plaisant et al. 1998	Unstructured	Text	No	Text	Time	No	Extraction visualization	Extractive Abstractive
Tufte 1994	Structured Unstructured	Document	No	Document	No	No	Extraction	Extractive
Rogers 1979	Structured	Document	No	Text	No	No	Extraction	Extractive Abstractive

### 5.3. Discussion

In order to clearly address the challenges related to EHR summarization we compared and discussed the existing studies based on the following criteria:

- Type of input Data (C1) this criterion refers to the input data which could be: structured data, semi-structured data and unstructured data.
- Data type (C2) : this criterion describe the type of data incorporated (text, xml, numeric, video, image)
- Representation standard (C3): this criterion indicates if the approach incorporates standard(i.e. information based standard, document based standard or Hybrid standard) (e.g., Yes or No).
- Summarization technique (C4): this criterion refers to the techniques deployed to summarize EHR which could be: grouping, compression, analysis, pattern- mining, classification, visualization
- Summarization approach(C5): this criterion refers to the target of the summarization approach abstractive , extractive , informative or indicative.
- medical knowledge based summarization(C6): this criterion describe the medical knowledge that system incorporate (e.g , Yes or No ).
- Output type (C7): this criterion concerns the type of displayed summarized data which is a combination of: numerical data, textual data, document, graph.
- Context-aware criterion (C8) this criterion refers to the context of data or the device and we defined: (i) Partial, used to demonstrate if an existing system uses concepts about the deployed context of the devices (e.g., time, location, and trajectory) or concepts about the static data and (ii) Total, used to determine if an existing system uses both of deployed context of devices and other static data context.
- User oriented summarization (C9) :this criterion represent that the approach oriented user (e.g., yes or No).

Our analysis summarized in Table 5 demonstrates that approaches are based on either structured data [103],[72],[7] or unstructured data [90],[103],[7],[89],[10],[93] and none of them rely on both structured and unstructured data in order to construct the summary. Another important part of this study is the output type of summarized data. The reviewed systems are either document-based [74],[102],[78] or graph based [88] They do not propose dedicated tools that make the summary accessible to the user nor provide them with appropriate perceptions of their needs. Users are more and more concerned about security, confidentiality, understanding their data, and the accuracy and completeness of their data. This should be mandatory when developing a visualization method for the summary to empower them with easier means. An intuitive and friendly GUI will benefit summarization-based systems. Also, none of the studies surveyed in this paper is user oriented and satisfy various users' needs. Our comparison highlights that the

evolution of the summary is still an open challenge. So, we observe that most of existing studies do not consider the context of the data in their analysis and do not consider the context on creating the summary and they rely only on the time property except some of them considering the time in their analysis [78],[74],[88]. Thus, existing systems are still unable to contextually interpret and reason on the transferred knowledge among real data, and consequently cannot synthesize data in order to provide accurate desired results. All existing systems focus on one objective, while none of them provide in the same framework various functionalities despite its importance in supporting users' preferences to find the data according to various needs. All objectives should be an integral part of a summarization-based system. Finally, we deduce that most of the studied approaches are extractive-based [73],[97],[72],[73],[91],[74],[91]. three of them [7],[80],[81] are abstractive-based and five studies [100],[95],[80],[10]. are extractive and abstractive-based approaches. From this comparative study, we deduce main 4 limitations:

1. **The lack of access and collect data from Medical Devices** Due to the heterogeneity of applications it seems critical to synthesize health data in order to provide a relevant, comprehensive and understanding view of the patient's history to effectively help clinical diagnostics.
2. **The lack of semantic interoperability** Applications generate a huge amount of heterogeneous data which makes it almost difficult to synthesize the knowledge communicating between clinical applications and provide efficient results.
3. **The lack of linking Data and medical devices to their contexts** We have to deconvolve the data and the device context towards identifying its capacity and its reliability to ensure the data consistency of the gathered data and to easily repair it when necessary.
4. **The lack of user-centered summary design** Increased cognitive workload of clinician has consistently been linked to the Text summarization. This make it nearly impossible to provide interactive and personalized summary. Existing systems are unable to generate adaptive summaries adjusting based on clinician preference and needs.

## 6. Electronic Health Record Visualization

### 6.1. Electronic Health Record Visualization definition

[104] has defined "Computer-based visualization as a visual system helping users to carry out some activities more effectively and efficiently by providing various visual representations of data sets. We distinguish two types of visualization systems: (i) Information visualization focusing on data sets with non spatial data attributes and discrete observations[105] and (ii) Scientific visualization visualize real objects in a spatial three dimensional-space [106]. In [107] authors shows that the Electronic Health Record have to combine heterogeneous visualization techniques to model relevant data from both information visualization and scientific visualization. Information visualization aims to map collected data to compact representations to generate meaningful and relevant information quickly. So, we conclude that information visualization is crucial to explore and query the heterogeneous and temporal data. On the other

hand, interactive information visualization ensures analysis exploration [108] supported by human cognition and visual abilities.

## 6.2. Electronic Health Record visualization approaches

In this section, we reviewed 40 papers focusing on visualization and visual analytics of EHR data and compared them according to the same criteria used to compare EHR summarization approaches. [104],[100] and [109] provide and discuss existing studies of EHR Visualisation and provide a new visualization tool called, LetterVis. The proposed tool support the analysis of clinic letters using interactive visual designs and queries. The authors model a common letter and incorporate in the same interface so to explore the related content and patterns. The text is analyzed and processed using Natural Language processing (NLP) technique and explored in multiple linked interactive views. mcNabb2019multivariate surveys work by focusing on healthcare visualization and visual analytics. [110] proposes a new approach based on visual-interactive that train models for prostate cancer identification. In [4] authors compare phenotype through a visual analysis tool named PhenoBlocks and defined a new hierarchy comparison algorithm. [111] proposes a novel algorithm aiming at simplifying the topology by eliminating duplicates, and incorporating natural language queries for searching anomalies. [4] described a new tool named PhenoLines visualizing a temporal evolution of phenotypes. In [112] authors provide a novel approach extracting heterogeneous data to an hierarchical task abstraction. Authors also proposed a new visualization tool called IDMVis to sequence temporal multidimensional events with their correlated data. [113] aims at increasing interpretability and interactivity of RNNs through visual analytic tool based on health experts, artificial intelligence scientists, and visual analytics researchers. In [114], authors aims at modeling semantically meaningful care plans and their processes and identifying critical events which help adapt these plans. The proposed approach defined an unsupervised algorithm following three key steps: (i) estimating event representation, (ii) warping and alignment event sequence and (iii) segmenting the sequences. Authors also suggest a novel visualization system illustrating the results. [115] presents PatientExplore producing an interactive and dynamic interfaces to facilitate the visualization of medical data [110]. Authors provided a new tool segmenting patients' data and aggregating them using the whole history and the combination of the treatments. Finally, it visualizes the aggregated data in an interactive way. In [116], authors provide a new framework named VIVID to visualize the missing values by exploring, identifying and validating their imputations. [117] provides a new tool for interactive processing of clinical notes through multiple NLP models [118].

The above study shows the importance of Natural Language Processing tools is analyzing and visualizing clinical text data using especially the open source GATE tool supporting text mining of biomedical documents [119]. However, authors demonstrate that it is unable to perform advanced analysis since it lacks interactive visualization features. [112] proposes a new visual analytic tool named AnamneVis incorporating NLP algorithms to extract structured medical information from both the dialogue between the clinician and his patient and the textual reports. [14] authors follow LifeLines2 approach [11] to extract structured data hidden inside clinical letters through a visual representation and using NLP method for the event system simplification technique. [120] proposes a new tool named LifeFlow used to visualize



interactively the event sequence data. In [121] authors provide a new tool aiming at simplifying temporal event sequence data, aggregating data and identifying hidden trends. [122] provides a new visualization tool named TimeSpan aiming at exploring the temporal aspects of a care plan process.

[101] describes EventThred, a visualization system evaluating some defined patterns in event sequence data. It measures the importance of the text in the document using the Term Frequency - Inverse Document Frequency (TF-IDF) [123]. [124] defined the Community Health Map, a new tool aiming at visualizing public healthcare datasets. It allows clinicians to visualize heterogeneous datasets gathered from Hospitals. In [123], authors propose a SIMID, a new tool for infectious diseases surveillance and spatio-temporal visualization. SIMID learns existing clinical data and simulates the spread of infectious disease. [101] introduces a new visualization tool named DICON. The proposed tool explores the similarity in cohorts of patients and generates new clusters.

Table 5: Summary of retrieved studies on electronic health record visualisation

Approach	Input Data	Data Type	Standard	Output	Context	User Oriented	Visualization Technique	Medical Knowledge
wang et al 2021	Unstructured	Text	No	Text	No	Yes	NLP	yes
Kwon et al 2020	Unstructured	Not mentioned	No	Graphical	No	Yes	ML	Yes
Jin et al 2020	Unstructured	Not mentioned	No	Graphical	Time	Yes	ESS	No
MCNaBB and Laramée 2019	Unstructured	Not mentioned	No	Map	No	No	Geospatial Visualization	No
GUO et al 2019	Structured Unstructured	Not mentioned	No	Pattern	No	Yes	Event Sequence Simplification	No
GlickSberg et al 2019	Unstructured	Not mentioned	Yes	EHR	Time	Yes	Clustering	No
Bernard et al 2019	Structured Unstructured	Document	Yes	EHR	Time	Yes	ESS	Yes
Almezaadeh et al 2019	Structured	Numerical	No	Network	Not mentioned	Yes	Geospatial	No
Trivedi et al 2018	Unstructured	Text	No	Text	No	No	NLP	No
Tong et al 2018	Structured	Geographic data	Yes	Cartogram	No	No	Analysis Classification	
Guo et al 2018	Unstructured	Numerical	Yes	Ontology	No	Yes	ESS	No
Ola and Sedig 2016	Unstructured	Text	Yes	Pattern	Yes	Yes	Gesopastial visualization	No
Loorak et al 2016	Structured Unstructured	Text	No	Temporal data	No	Yes	ESS	YES
kamaleswaran et al 2016	Unstructured	Numerical	No	Temporal data	Time	No	clustering	Yes
Jiang et al 2016	Unstructured	Text	Yes	Geospatial information	No	Yes	Geospatial visualization	No
Glueck et al 2016	Unstructured	Text	No	Ontology	Yes	Yes	Comparison	yes

Bernard et al 2014	Unstructured	Not mentioned	No	Temporal attribute	No	No	No	ML	Yes
Malik et al 2014	Structured	Database	No	Numerical	Time	Yes	ESS	ESS	Yes
Kamaleswaran et al 2014	Unstructured	Stream data	No	Stream	Time	Yes	Clustering	Clustering	Yes
GHOTOZ et al 2014	Unstructured	Numerical	No	Matrix	No	Yes	ESS	ESS	Yes
Borland et al 2014	Unstructured	Numerical	No	Curve	time	Yes	Comparison	Comparison	Yes
Monore et al 2013	Structured	Numerical	No	Pattern	Time	No	ESS	ESS	No
wong and Ghotz 2012	Structured	Not mentioned	No	Pattern	Time	Yes	ESS	ESS	Yes
Sopan et al 2012	Unstructured	Not mentioned	No	Pattern	Time	Yes	ESS	ESS	Yes
Alonso et al 2012	Unstructured	Text	No	Text	No	No	Extraction	Extraction	No
Gschwandter et al 2011	Unstructured	Not mentioned	No	Numerical	Time	Yes	ESS	ESS	Yes
wang et al 2009	Structured	Not mentioned	No	temporal data	Time	Yes	Clustering	Clustering	Yes
BUI et al 2007	Structured	Xml	No	Document	Time	Yes	Classification	Classification	Yes
Fails et al 2006	Unstructured	Numerical	No	Pattern	Time	Yes	ESS	ESS	Yes
Hinum et al 2005	Unstructured	Temporal data	No	Curve, Graphics	Time	No	Classification	Classification	Yes
Goren et al 2004	Structured	Not mentioned	No	Graphics	Time	Yes	Clustering	Clustering	Yes
Bade et al 2004	Unstructured	Not mentioned	No	Graphics	Time	No	Abstraction	Abstraction	Yes
Horan et al 2001	Unstructured	Text	No	No	Time	No	Extraction	Extraction	Yes
Plaisant et al 1998	Unstructured	Document	No	Document	No	No	Extraction	Extraction	No

### 6.3. Discussion

Our analysis summarized in Table 5 demonstrates that approaches are based on either structured data [120],[125],[126],[126],[112], [? ],[87],[80] Only two approaches rely on both structured and unstructured data [44],[122]. Another important part of this study is the output type in the visualization. Most of the reviewed systems are Pattern-based [44],[127],[121],[87],[107] graphical-based [125],[126],[86],[128] or text-based [120],[129],[127],[122]. They do not propose dedicated tools that make the visualization accessible to the user nor provide them with appropriate perceptions of their needs. Also, all the studies surveyed in this paper are user oriented and satisfy various users' needs. Our comparison highlights that the evolution of the summary is still an open challenge. So, we observe that most of existing studies [87],[125],[130], do not consider the context of the data in the visualization process except some of them [126],[115],[80],[121],[131].consider the time in their analysis. From this comparative study, we deduce some major challenges faced in Electronic Health Record data visualization:

1. **Providing an accessible Electronic Health Record** providing an accessible Electronic Health Record is very challenging [132]. Many challenges have to be addressed related to relevant data acquisition and data extraction since it requires amount of time to search for. Clinical data is very sensitive and are often unstructured. It seems important to convert the data into a structured form but it risks losing valuable insight. It also seems crucial to apply an anonymization process on Electronic Health Record data taking into account the data governance group policies.
2. **Ensuring data quality** Data can contain incomplete and erroneous values since it is entered and computed manually. For that, it seems crucial to verify data quality. In the other hand, we note that EHRs contains very heterogenous data and do not create with supporting research in mind [132]. However, amount of research about EHR are emerged and widely accepted worldwide, which improves the quality control measures for collecting clinical data [133].
3. **Improving visualization techniques**, in the literature we found five visualisation-based techniques such as machine learning, natural language , event sequence simplification, geospatial visualization, clustering and comparison. Firstly, we note the importance of incorporating Machine Learning (ML) techniques into EHR Visualization [125],[134]. We highlight that it is a new field of research but combined with EHR visualization it is not very mature yet. EHR visualization research can benefit more with the help of ML techniques by describing existing states, predicting health problems and providing guidance for the future based on domain knowledge. Many ML techniques are applied in EHR visual analytic to increase automation of processing, specially, deep learning [135], neural networks[125], support vector machines [112] and topic models [109]. Secondly, EHRs contains diverse and heterogeneous data requiring appropriate modifiers to capture words, phrases and their relationships. Thus, we note that NLP techniques can be efficient in transforming unstructured text into structured data [87]. EHRs includes temporal data and events and it seems important to model and analyze these temporal elements by improving both data-processing and visualization of EHRs. In this context, the Event Sequence

Simplification technique to reduce the visual complexity of event sequences in aggregated model [121]. LifeLines [90] and EventFlow [121] adopted such techniques. The fourth technique used in visualization process is geospatial visualization so the most important work for this technique [123] introduces SIMID, a new spatio-temporal visualization tool aiming at surveilling infectious diseases. The proposed system uses interactive animated maps to analyse the infectious disease. Finally, another technique seems critical to provide efficient EHR analysis and to develop information visualization discipline. This shows our comparison table ?? where some approaches [136],[87] produce homogeneous subgroups based on similarities by the use of hierarchical clustering algorithms

4. **Making data interoperable** our study confirms the lack of standard definition of an EHR. Providers often provide specific standard to support specific clinical domain and care process [137]. We highlight also that the use of terminologies such as the UMLS combined with health standards should making data more interoperable.

Many studies are proposed in order to address the above challenges such as building a freely accessibility EHR database [109] and improving data validation and interoperability [107].

## 7. conclusion

Health Information Systems have to obey the inetreoperability systems in order to efficiently manage and deliver relevant clinical data and ensure better communication between individuals, processes and technology. Being interoperable ensuring healthcare services improvement, the increase and the efficiency of their quality. Although the summarization and the visualization process are essential for clinicians to provide coordinated care and practice effective communication. Also, heterogeneous clinical systems are implemented in health care either in the EHR summarization or EHR visualization. In this study we focused on interoperability challenge and the use of health standards and terminologies to ensure the interoperability specially in the summarization-based EHR systems and visualization-based EHR systems. Here, we described, compared and discussed approaches related to interoperability such as e-health standards, terminologies and IoT ontologie. We also reviewed and discussed summarization-based EHR systems and visualization-based EHR systems in order to schown how interoperability issues can enhance EHR analysis to build accurate summarization-based EHR systems and visualization-based EHR systems Finally, our discussions reveal that five techniques should be adopted to enhance inetreoperability and so building free accessibility and accurate EHR and improving data validation and interoperability.

## References

- [1] J. A. Blaya, H. S. Fraser, B. Holt, E-health technologies show promise in developing countries, *Health Affairs* 29 (2010) 244–251.
- [2] F. Adebesin, P. Kotzé, D. Van Greunen, R. Foster, Barriers & challenges to the adoption of e-health standards in africa (2013).

- [3] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, P. Group\*, Preferred reporting items for systematic reviews and meta-analyses: the prisma statement, *Annals of internal medicine* 151 (2009) 264–269.
- [4] M. Glueck, M. P. Naeini, F. Doshi-Velez, F. Chevalier, A. Khan, D. Wigdor, M. Brudno, Phenolines: Phenotype comparison visualizations for disease subtyping via topic models, *IEEE Transactions on Visualization and Computer Graphics* 24 (2017) 371–381.
- [5] L. Paladin, L. Hirsh, D. Piovesan, M. A. Andrade-Navarro, A. V. Kajava, S. C. Tosatto, Repeatsdb 2.0: improved annotation, classification, search and visualization of repeat protein structures, *Nucleic acids research* 45 (2017) D308–D312.
- [6] H. Oh, C. Rizo, M. Enkin, A. Jadad, What is ehealth (3): a systematic review of published definitions, *Journal of medical Internet research* 7 (2005) e110.
- [7] S. M. Powsner, E. R. Tufte, Summarizing clinical psychiatric data, *Psychiatric Services* 48 (1997) 1458–1460.
- [8] R. Pivovarov, A. J. Perotte, E. Grave, J. Angiolillo, C. H. Wiggins, N. Elhadad, Learning probabilistic phenotypes from heterogeneous ehr data, *Journal of biomedical informatics* 58 (2015) 156–165.
- [9] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, M. Zhou, Low-quality product review detection in opinion summarization, in: *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007, pp. 334–342.
- [10] Q. E. Whiting-O’Keefe, D. W. Simborg, W. V. Epstein, A controlled experiment to evaluate the use of a time-oriented summary medical record, *Medical care* (1980) 842–852.
- [11] Q. Wang, R. S. Laramée, A. Lacey, W. O. Pickrell, Lettervis: a letter-space view of clinic letters, *The Visual Computer* 37 (2021) 2643–2656.
- [12] D. G. Altman, K. F. Schulz, D. Moher, M. Egger, F. Davidoff, D. Elbourne, P. C. Gøtzsche, T. Lang, C. Group, The revised consort statement for reporting randomized trials: explanation and elaboration, *Annals of internal medicine* 134 (2001) 663–694.
- [13] J. P. Vandenbroucke, E. Von Elm, D. G. Altman, P. C. Gøtzsche, C. D. Mulrow, S. J. Pocock, C. Poole, J. J. Schlesselman, M. Egger, S. Initiative, Strengthening the reporting of observational studies in epidemiology (strobe): explanation and elaboration, *PLoS medicine* 4 (2007) e297.
- [14] I. Berges, J. Bermúdez, A. Illarramendi, Toward semantic interoperability of electronic health records, *IEEE Transactions on Information Technology in Biomedicine* 16 (2011) 424–431.
- [15] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, G. B. Laleci, A survey and analysis of electronic healthcare record standards, *Acm Computing Surveys (Csur)* 37 (2005) 277–315.
- [16] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D. W. Bates, B. Middleton, The value of health care information exchange and interoperability: There is a business case to be made for spending money on a fully standardized nationwide system., *Health affairs* 24 (2005) W5–10.
- [17] N. Denjoy, European coordination committee of the radiological, electromedical, and healthcare it industry: Medical imaging equipment age, profile & density, 2016.
- [18] L. E. Whitman, H. Panetto, The missing link: Culture and language barriers to interoper-

- ability, *Annual Reviews in Control* 30 (2006) 233–241.
- [19] H. Van Der Veer, A. Wiles, Achieving technical interoperability, European telecommunications standards institute (2008).
  - [20] J. R. Hobbs, F. Pan, An ontology of time for the semantic web, *ACM Transactions on Asian Language Information Processing (TALIP)* 3 (2004) 66–85.
  - [21] O. Bodenreider, The unified medical language system (umls): integrating biomedical terminology, *Nucleic acids research* 32 (2004) D267–D270.
  - [22] C. Europea, Ict standards in the health sector: current situation and prospect, *Special Study n 1* (2008).
  - [23] B. Blobel, C. González, F. Oemig, D. M. Lopez, P. Nykänen, P. Ruotsalainen, The role of architecture and ontology for interoperability., in: *EFMI-STC*, 2010, pp. 33–39.
  - [24] P. Hirmer, M. Wieland, U. Breitenbücher, B. Mitschang, Dynamic ontology-based sensor binding, in: *East European Conference on Advances in Databases and Information Systems*, Springer, 2016, pp. 323–337.
  - [25] L. Daniele, F. d. Hartog, J. Roes, Created in close interaction with the industry: the smart appliances reference (saref) ontology, in: *International Workshop Formal Ontologies Meet Industries*, Springer, 2015, pp. 100–112.
  - [26] D. J. Russomanno, C. Kothari, O. Thomas, Sensor ontologies: from shallow to deep models, in: *Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory*, 2005. SSST’05., IEEE, 2005, pp. 107–112.
  - [27] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, D. Hong, A. Johansen, J. Koh, J. Ploennigs, Y. Agarwal, et al., Brick: Towards a unified metadata schema for buildings, in: *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, 2016, pp. 41–50.
  - [28] M. Haghi, K. Thurow, R. Stoll, Wearable devices in medical internet of things: scientific research and commercially available devices, *Healthcare informatics research* 23 (2017) 4–15.
  - [29] R. T. Fielding, Architectural styles and the design of network-based software architectures, University of California, Irvine, 2000.
  - [30] A. M. Ouksel, A. Sheth, Semantic interoperability in global information systems, *ACM Sigmod Record* 28 (1999) 5–12.
  - [31] I. ISO, 21090 health informatics-harmonized data types for information interchange, Bericht, ISO/CEN (2011).
  - [32] I. ISO, Guide 2–standardization and related activities-general vocabulary, Geneva: ISO (2004).
  - [33] D. Kalra, T. Beale, S. Heard, The openehr foundation, *Studies in health technology and informatics* 115 (2005) 153–173.
  - [34] F. Adebesin, P. Kotze, R. Foster, D. Van Greunen, A review of interoperability standards in e-health and imperatives for their adoption in africa, *South African Computer Journal* 50 (2013) 55–72.
  - [35] P. Aspden, J. M. Corrigan, J. Wolcott, S. M. Erickson, Committee on data standards for patient safety, Patient safety: achieving a new standard for care [cited 2005 Oct 3]. Available from: URL: <http://www.iom.edu/report.asp> (2004).



- [36] G. Dickinson, L. Fischetti, S. Heard, HL7 ehr system functional model draft standard for trial use, Health Level 7 (2004).
- [37] A. L. Rector, The interface between information, terminology, and inference models, in: MEDINFO 2001, IOS Press, 2001, pp. 246–250.
- [38] A. Geraci, IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries, IEEE Press, 1991.
- [39] S. Garde, P. Knaup, E. J. Hovenga, S. Heard, Towards semantic interoperability for electronic health records, *Methods of information in medicine* 46 (2007) 332–343.
- [40] S. Heiler, Semantic interoperability, *ACM Computing Surveys (CSUR)* 27 (1995) 271–273.
- [41] T. Benson, Principles of health interoperability HL7 and SNOMED, Springer Science & Business Media, 2012.
- [42] N. E. M. Association, et al., Digital imaging and communications in medicine (dicom) part 1: Introduction and overview.(2011), NEMA: Virginia (2011).
- [43] S. E. GROUP, et al., Sdmx-statistical data and metadata exchange, 2006.
- [44] J. Guo, K. Araki, K. Tanaka, J. Sato, M. Suzuki, A. Takada, T. Suzuki, Y. Nakashima, H. Yoshihara, The latest mml (medical markup language) version 2.3—xml-based standard for medical data exchange/storage, *Journal of Medical Systems* 27 (2003) 357–366.
- [45] J.-W. Kim, S.-H. Jeon, C.-M. Lim, S.-Y. Park, N.-H. Kim, Implementation of reporting system for continuity of care document based on web service, in: *Proceedings of the IEEK Conference, The Institute of Electronics and Information Engineers*, 2009, pp. 402–404.
- [46] K. W. Boone, The CDA tm book, Springer Science & Business Media, 2011.
- [47] C. J. McDonald, S. M. Huff, J. G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, J. Hook, et al., Loinc, a universal standard for identifying laboratory observations: a 5-year update, *Clinical chemistry* 49 (2003) 624–633.
- [48] P. Tyrer, M. Crawford, R. Mulder, R. Blashfield, A. Farnam, A. Fossati, Y.-R. Kim, N. Koldobsky, D. Lecic-Tosevski, D. Ndeti, et al., The rationale for the reclassification of personality disorder in the 11th revision of the international classification of diseases (icd-11), *Personality and Mental Health* 5 (2011) 246–259.
- [49] W. H. Organization, et al., International classification of primary care, -icpc-2 (2009).
- [50] A. M. Association, et al., American medical association current procedural terminology cpt 2017, 2017.
- [51] H.-a. Park, N. Hardiker, Clinical terminologies: a solution for semantic interoperability, *Journal of Korean Society of Medical Informatics* 15 (2009) 1–11.
- [52] R. Alterman, Understanding and summarization, *Artificial Intelligence Review* 5 (1991) 239–254.
- [53] D. J. Brailer, Interoperability: The key to the future health care system: Interoperability will bind together a wide network of real-time, life-critical data that not only transform but become health care., *Health affairs* 24 (2005) W5–19.
- [54] C. N. Mead, et al., Data interchange standards in healthcare it-computable semantic interoperability: Now possible but still difficult. do we really need a better mousetrap?, *Journal of Healthcare Information Management* 20 (2006) 71.
- [55] K. A. Wager, F. W. Lee, J. P. Glaser, Managing health care information systems: a practical approach for health care executives, John Wiley & Sons, 2005.

- [56] H. Riazi, M. Jafarpour, E. Bitaraf, Towards national ehealth implementation-a comparative study on who/itu national ehealth strategy toolkit in iran., in: MIE, 2014, pp. 246–250.
- [57] D. Bender, K. Sartipi, HL7 fhir: An agile and restful approach to healthcare information exchange, in: Proceedings of the 26th IEEE international symposium on computer-based medical systems, IEEE, 2013, pp. 326–331.
- [58] R. Hussein, U. Engelmann, A. Schroeter, H.-P. Meinzer, Dicom structured reporting: Part 1. overview and characteristics, Radiographics 24 (2004) 891–896.
- [59] M. Q. Stearns, C. Price, K. A. Spackman, A. Y. Wang, Snomed clinical terms: overview of the development process and project status., in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2001, p. 662.
- [60] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. Shabo, HL7 clinical document architecture, release 2, Journal of the American Medical Informatics Association 13 (2006) 30–39.
- [61] T. Tudorache, C. I. Nyulas, N. F. Noy, M. A. Musen, Using semantic web in icd-11: three years down the road, in: International Semantic Web Conference, Springer, 2013, pp. 195–211.
- [62] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al., The ssn ontology of the w3c semantic sensor network incubator group, Journal of Web Semantics 17 (2012) 25–32.
- [63] L. Xue, Y. Liu, P. Zeng, H. Yu, Z. Shi, An ontology based scheme for sensor description in context awareness system, in: 2015 IEEE International Conference on Information and Automation, IEEE, 2015, pp. 817–820.
- [64] T. Flury, G. Privat, F. Ramparany, Owl-based location ontology for context-aware services, Proceedings of the Artificial Intelligence in Mobile Systems (AIMS 2004) (2004) 52–57.
- [65] R. Fikes, Q. Zhou, A reusable time ontology, in: AAAI-2002 Workshop on Ontologies and the Semantic Web, Citeseer, 2002.
- [66] M. Bermudez-Edo, T. Elsaleh, P. Barnaghi, K. Taylor, Iot-lite: a lightweight semantic model for the internet of things, in: 2016 INTL IEEE conferences on ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (uic/atc/scalcom/cbdcom/iop/smartworld), IEEE, 2016, pp. 90–97.
- [67] A. Dridi, S. Sassi, R. Chbeir, S. Faiz, A flexible semantic integration framework for fully-integrated ehr based on fhir standard., in: ICAART (2), 2020, pp. 684–691.
- [68] R. Miotto, L. Li, J. T. Dudley, Deep learning to predict patient future diseases from the electronic health records, in: European conference on information retrieval, Springer, 2016, pp. 768–774.
- [69] I. Iakovidis, Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in europe, International journal of medical informatics 52 (1998) 105–115.
- [70] T. D. Gunter, N. P. Terry, The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions, Journal of medical Internet research 7 (2005) e383.
- [71] D. W. Bates, M. Ebell, E. Gotlieb, J. Zapp, H. Mullins, A proposal for electronic medical

- records in us primary care, *Journal of the American Medical Informatics Association* 10 (2003) 1–10.
- [72] J. L. Rogers, O. M. Haring, R. A. Watson, Automating the medical record: emerging issues, in: *Proceedings of the Annual Symposium on Computer Application in Medical Care*, American Medical Informatics Association, 1979, p. 255.
  - [73] D. J. McInerney, B. Dabiri, A.-S. Touret, G. Young, J.-W. Meent, B. C. Wallace, Query-focused ehr summarization to aid imaging diagnosis, in: *Machine Learning for Healthcare Conference*, PMLR, 2020, pp. 632–659.
  - [74] J. J. Gong, J. V. Gutttag, Learning to summarize electronic health records using cross-modality correspondences, in: *Machine learning for healthcare conference*, PMLR, 2018, pp. 551–570.
  - [75] Z. O. Abu-Faraj, S. S. Barakat, M. H. Chaleby, J. D. Zaklit, A sim card-based ubiquitous medical record bracelet/pendant system—a pilot study, in: *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, volume 4, IEEE, 2011, pp. 1914–1918.
  - [76] L. Tapak, H. Mahjub, O. Hamidi, J. Poorolajal, Real-data comparison of data mining methods in prediction of diabetes in iran, *Healthcare informatics research* 19 (2013) 177–185.
  - [77] E. Soysal, J. L. Warner, J. C. Denny, H. Xu, Identifying metastases-related information from pathology reports of lung cancer patients, *AMIA Summits on Translational Science Proceedings* 2017 (2017) 268.
  - [78] M. Moradi, N. Ghadiri, Different approaches for identifying important concepts in probabilistic biomedical text summarization, *Artificial intelligence in medicine* 84 (2018) 101–116.
  - [79] J. S. Hirsch, J. S. Tanenbaum, S. Lipsky Gorman, C. Liu, E. Schmitz, D. Hashorva, A. Ervits, D. Vawdrey, M. Sturm, N. Elhadad, Harvest, a longitudinal patient record summarizer, *Journal of the American Medical Informatics Association* 22 (2015) 263–274.
  - [80] A. A. Bui, D. R. Aberle, H. Kangarloo, Timeline: visualizing integrated patient records, *IEEE Transactions on Information Technology in Biomedicine* 11 (2007) 462–473.
  - [81] V. Bashyam, W. Hsu, E. Watt, A. A. Bui, H. Kangarloo, R. K. Taira, Problem-centric organization and visualization of patient imaging and clinical data, *Radiographics* 29 (2009) 331–343.
  - [82] Y. Shahar, D. Goren-Bar, D. Boaz, G. Tahan, Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions, *Artificial intelligence in medicine* 38 (2006) 115–135.
  - [83] J. Hunter, Y. Freer, A. Gatt, R. Logie, N. McIntosh, M. Van Der Meulen, F. Portet, E. Reiter, S. Sripada, C. Sykes, Summarising complex icu data in natural language, in: *Amia annual symposium proceedings*, volume 2008, American Medical Informatics Association, 2008, p. 323.
  - [84] P. D. Clayton, S. P. Narus, W. A. Bowes III, T. S. Madsen, A. B. Wilcox, G. Orsmond, B. Rocha, S. N. Thornton, S. Jones, C. A. Jacobsen, et al., Physician use of electronic medical records: issues and successes with direct data entry and physician productivity, in: *AMIA annual symposium proceedings*, volume 2005, American Medical Informatics

Association, 2005, p. 141.

- [85] M. E. Matheny, R. A. Miller, T. A. Ikizler, L. R. Waitman, J. C. Denny, J. S. Schildcrout, R. S. Dittus, J. F. Peterson, Development of inpatient risk stratification models of acute kidney injury for use in electronic health records, *Medical Decision Making* 30 (2010) 639–650.
- [86] R. Bade, S. Schlechtweg, S. Miksch, Connecting time-oriented data and information to a coherent interactive visualization, in: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 105–112.
- [87] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, M. Smith, Temporal summaries: Supporting temporal categorical searching, aggregation and comparison, *IEEE transactions on visualization and computer graphics* 15 (2009) 1049–1056.
- [88] D. Borland, V. L. West, W. E. Hammond, Multivariate visualization of system-wide national health service data using radial coordinates, in: *Proc. Workshop on Visual Analytics in Healthcare*, 2014.
- [89] A. Carenini, D. Cerri, R. Krummenacher, E. Simperl, Enabling interoperability of patient summaries across europe with triplespaces, in: *Interoperability in Healthcare Information Systems: Standards, Management, and Technology*, IGI Global, 2013, pp. 232–249.
- [90] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, B. Shneiderman, Lifelines: using visualization to enhance navigation and analysis of patient records, in: *The craft of information visualization*, Elsevier, 2003, pp. 308–312.
- [91] H. Liu, C. Friedman, Cliniviewer: a tool for viewing electronic medical records based on natural language processing and xml, in: *MEDINFO 2004*, IOS Press, 2004, pp. 639–643.
- [92] A. Wright, J. Pang, J. C. Feblowitz, F. L. Maloney, A. R. Wilcox, H. Z. Ramelson, L. I. Schneider, D. W. Bates, A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record, *Journal of the American Medical Informatics Association* 18 (2011) 859–867.
- [93] D. Radev, E. Hovy, K. McKeown, Introduction to the special issue on summarization, *Computational linguistics* 28 (2002) 399–408.
- [94] A. Nenkova, K. McKeown, A survey of text summarization techniques, in: *Mining text data*, Springer, 2012, pp. 43–76.
- [95] P. S. Lukas, R. Krummenacher, F. D. Biasiutti, S. Bégé, H. Znoj, R. von Känel, Association of fatigue and psychological distress with quality of life in patients with a previous venous thromboembolic event, *Thrombosis and haemostasis* 102 (2009) 1219–1226.
- [96] L. H. Reeve, H. Han, A. D. Brooks, The use of domain-specific concepts in biomedical text summarization, *Information Processing & Management* 43 (2007) 1765–1776.
- [97] J. Liang, C.-H. Tsou, A. Poddar, A novel system for extractive clinical note summarization using ehr data, in: *Proceedings of the 2nd clinical natural language processing workshop*, 2019, pp. 46–54.
- [98] A. S. Razavian, J. Sullivan, S. Carlsson, A. Maki, Visual instance retrieval with deep convolutional networks, *ITE Transactions on Media Technology and Applications* 4 (2016) 251–258.
- [99] Y. Handelsman, Z. T. Bloomgarden, G. Grunberger, G. Umpierrez, R. S. Zimmerman, T. S. Bailey, L. Blonde, G. A. Bray, A. J. Cohen, S. Dagogo-Jack, et al., American association

of clinical endocrinologists and american college of endocrinology—clinical practice guidelines for developing a diabetes mellitus comprehensive care plan—2015—executive summary, *Endocrine Practice* 21 (2015) 413–437.

- [100] V. L. West, D. Borland, W. E. Hammond, Innovative information visualization of electronic health record data: a systematic review, *Journal of the American Medical Informatics Association* 22 (2015) 330–339.
- [101] D. Gotz, H. Stavropoulos, Decisionflow: Visual analytics for high-dimensional temporal event sequence data, *IEEE transactions on visualization and computer graphics* 20 (2014) 1783–1792.
- [102] C. Xiao, E. Choi, J. Sun, Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review, *Journal of the American Medical Informatics Association* 25 (2018) 1419–1428.
- [103] D. Nallaperuma, D. De Silva, et al., A participatory model for multi-document health information summarisation, *Australasian Journal of Information Systems* 21 (2017).
- [104] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, Interactive information visualization to explore and query electronic health records, *Foundations and Trends in Human-Computer Interaction* 5 (2013) 207–298.
- [105] M. Tory, T. Moller, Rethinking visualization: A high-level taxonomy, in: *IEEE symposium on information visualization*, IEEE, 2004, pp. 151–158.
- [106] B. Preim, D. Bartz, *Visualization in medicine: theory, algorithms, and applications*, Elsevier, 2007.
- [107] J. A. Fails, A. Karlson, L. Shahamat, B. Shneiderman, A visual interface for multivariate temporal data: Finding patterns of events across multiple histories, in: *2006 IEEE Symposium On Visual Analytics Science And Technology*, IEEE, 2006, pp. 167–174.
- [108] G. Leinhardt, S. Leinhardt, Chapter 3: Exploratory data analysis: New tools for the analysis of empirical data, *Review of research in education* 8 (1980) 85–157.
- [109] S. Jiang, S. Fang, S. Bloomquist, J. Keiper, M. J. Palakal, Y. Xia, S. J. Grannis, Healthcare data visualization: Geospatial and temporal integration., in: *VISIGRAPP (2: IVAPP)*, 2016, pp. 214–221.
- [110] J. Bernard, D. Sessler, J. Kohlhammer, R. A. Ruddle, Using dashboard networks to visualize multiple patient histories: a design study on post-operative prostate cancer, *IEEE transactions on visualization and computer graphics* 25 (2018) 1615–1628.
- [111] M. Glueck, A. Gvozdk, F. Chevalier, A. Khan, M. Brudno, D. Wigdor, Phenostacks: cross-sectional cohort phenotype comparison visualizations, *IEEE transactions on visualization and computer graphics* 23 (2016) 191–200.
- [112] Y. Zhang, K. Chanana, C. Dunne, Idmvis: Temporal event sequence visualization for type 1 diabetes treatment decision support, *IEEE transactions on visualization and computer graphics* 25 (2018) 512–522.
- [113] B. C. Kwon, V. Anand, K. A. Severson, S. Ghosh, Z. Sun, B. I. Frohnert, M. Lundgren, K. Ng, Dpvis: Visual analytics with hidden markov models for disease progression pathways, *IEEE transactions on visualization and computer graphics* 27 (2020) 3685–3700.
- [114] S. Guo, Z. Jin, D. Gotz, F. Du, H. Zha, N. Cao, Visual progression analysis of event sequence data, *IEEE transactions on visualization and computer graphics* 25 (2018) 417–426.

- [115] B. S. Glicksberg, B. Oskotsky, P. M. Thangaraj, N. Giangreco, M. A. Badgeley, K. W. Johnson, D. Datta, V. A. Rudrapatna, N. Rappoport, M. M. Shervey, et al., Patientexplorer: an extensible application for dynamic visualization of patient clinical history from electronic health records in the omop common data model, *Bioinformatics* 35 (2019) 4515–4518.
- [116] S. Alemzadeh, U. Niemann, T. Ittermann, H. Völzke, D. Schneider, M. Spiliopoulou, K. Bühler, B. Preim, Visual analysis of missing values in longitudinal cohort study data, in: *Computer graphics forum*, volume 39, Wiley Online Library, 2020, pp. 63–75.
- [117] G. Trivedi, P. Pham, W. W. Chapman, R. Hwa, J. Wiebe, H. Hochheiser, Nlpreviz: an interactive tool for natural language processing on clinical text, *Journal of the American Medical Informatics Association* 25 (2018) 81–87.
- [118] T. A. Koleck, C. Dreisbach, P. E. Bourne, S. Bakken, Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review, *Journal of the American Medical Informatics Association* 26 (2019) 364–379.
- [119] V. Tablan, I. Roberts, H. Cunningham, K. Bontcheva, Gatecloud. net: a platform for large-scale, open-source text processing on the cloud, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013) 20120071.
- [120] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, B. Shneiderman, Lifeflow: visualizing an overview of event sequences, in: *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 1747–1756.
- [121] M. Monroe, R. Lan, H. Lee, C. Plaisant, B. Shneiderman, Temporal event sequence simplification, *IEEE transactions on visualization and computer graphics* 19 (2013) 2227–2236.
- [122] M. H. Loorak, C. Perin, N. Kamal, M. Hill, S. Carpendale, Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients, *IEEE transactions on visualization and computer graphics* 22 (2015) 409–418.
- [123] L. L. Ramírez-Ramírez, Y. R. Gel, M. Thompson, E. de Villa, M. McPherson, A new surveillance and spatio-temporal visualization tool simid: Simulation of infectious diseases using random networks and gis, *Computer methods and programs in biomedicine* 110 (2013) 455–470.
- [124] A. Sopan, A. S.-I. Noh, S. Karol, P. Rosenfeld, G. Lee, B. Shneiderman, Community health map: A geospatial and multivariate data visualization tool for public health datasets, *Government Information Quarterly* 29 (2012) 223–234.
- [125] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, J. Choo, Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records, *IEEE transactions on visualization and computer graphics* 25 (2018) 299–309.
- [126] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, N. Cao, Carepre: An intelligent clinical decision assistance system, *ACM Transactions on Computing for Healthcare* 1 (2020) 1–20.
- [127] O. Ola, K. Sedig, Beyond simple charts: Design of visualizations for big health data, *Online journal of public health informatics* 8 (2016).
- [128] W. Horn, C. Popow, L. Unterasinger, Support for fast comprehension of icu data: Visualization using metaphor graphics, *Methods of information in medicine* 40 (2001) 421–424.

- [129] C. Tong, L. McNabb, R. S. Laramee, Cartograms with topological features, in: *Proceedings of the Conference on Computer Graphics & Visual Computing*, 2018, pp. 127–134.
- [130] L. McNabb, R. S. Laramee, Multivariate maps—a glyph-placement algorithm to support multivariate geospatial visualization, *Information* 10 (2019) 302.
- [131] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, B. Shneiderman, An evaluation of visual analytics approaches to comparing cohorts of event sequences, in: *EHRVis Workshop on Visualizing Electronic Health Record Data at VIS*, volume 14, Citeseer, 2014.
- [132] M. Critical Data, *Secondary analysis of electronic health records*, Springer Nature, 2016.
- [133] E. Kim, S. M. Rubinstein, K. T. Nead, A. P. Wojcieszynski, P. E. Gabriel, J. L. Warner, The evolving use of electronic health records (ehr) for research, in: *Seminars in radiation oncology*, volume 29, Elsevier, 2019, pp. 354–361.
- [134] J. Bernard, D. Sessler, T. May, T. Schlomm, D. Pehrke, J. Kohlhammer, A visual-interactive system for prostate cancer cohort analysis, *IEEE computer graphics and applications* 35 (2015) 44–55.
- [135] R. Kamaleswaran, J. E. Pugh, A. Thommandram, A. James, C. McGregor, Visualizing neonatal spells: Temporal visual analytics of high frequency cardiorespiratory physiological event streams, in: *IEEE VIS 2014 Workshop on Visualization of Electronic Health Records*, 2014, pp. 1–4.
- [136] R. Xu, D. Wunsch, Survey of clustering algorithms, *IEEE Transactions on neural networks* 16 (2005) 645–678.
- [137] M. R. Cowie, J. I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay, et al., Electronic health records to facilitate clinical research, *Clinical Research in Cardiology* 106 (2017) 1–9.