Sensitive Data Discovery in Care Pathways using Business Process Modelling and HL7-CDA

Intidhar Essefi
University of Tunis el Manar, The Higher Institute of Medical Technologies of Tunis
Research Laboratory of Biophysics and Medical Technologies
Tunis, Tunisia
e-mail: essefi.intidhar@gmail.com

Hanen Boussi Rahmouni
University of the West of England, the Computer Science Research Center
Bristol, UK
e-mail: hanene4.rahmouni@uwe.ac.uk

Mohamed Fethi Ladeb
Radiology Department
Kassab Orthopedics Institute
Manouba, Tunisia
e-mail: fethiladeb@hotmail.fr

***Abstract*— Medical data communication is an important process enabling collaboration between healthcare professionals. The use of patient Electronic Health Record (EHR) ensures an enhanced continuity of care since it provides a centralized patient information access in a seamless way. In data protection law, the electronic exchange of medical data should comply with privacy obligations and data security safeguards. It is therefore a legal requirement for hospitals to ensure that patient information are processed and shared throughout clinical business processes in a standardized and structured form in order to be able to clearly discover and highlight the patient protected information manipulated in each process. In this work, we propose a clinical pathway specification methodology that is at the same time data driven and privacy aware. Our model gives special attention to the structure and the data content of shared medical documents. These documents are usually structured following the Health Level 7 (HL7)-Clinical Document Architecture (CDA). This research aims to define a legally shared HL7 structure of medical data ought to be processed and exchanged within clinical processes. Throughout the use of an ontology information model of medical data, we are aiming to capture and classify the data used in each clinical process into categories. We put special emphases on the level of protection required by each category of data in respect of the international health data legislation namely, the Health Insurance and Accountability Act (HIPAA). Particularly, our model ensures the implementation of the privacy by design principle since it ensures the adoption of data protection requirements starting from a very early stage of Hospital Information Systems (HIS) design.**

Keywords-business process modeling; clinical pathways; data driven; legally HL7 structured medical data; HIPAA legislation; ontology; patient privacy; privacy requirements; CDA.

I. INTRODUCTION

In the era of EHRs (Electronic Health Records), many questions are raised: who can access these records in the course of providing health care to the patient, under which circumstances and how? When, if so, can these documents be used for other purposes, such as public health, care quality improvement and public research? When is the consent of the data subject necessary for such "secondary" use of the patient's data? Without the consent of the patient, what are the implications for confidentiality and the sharing of personal medical records or their secondary use?

In order to improve services' quality within a hospital environment, it is important to automate the underlining workflows of each clinical process adopted within the hospital. This requires an explicit design and the implementation of business process models tailored for the concerned field. Medical healthcare is a multidisciplinary field. Its business processes and workflows are very complex. Throughout each medical process, several types of clinical information need to be circulated and treated within or without the hospital's boundaries. Medical data are produced, transmitted between medical departments and shared between healthcare professionals as specified by the clinical pathways enforced by the hospital information systems in use. Throughout this processes, several types of medical data documents are being manipulated, including admission papers, insurance documents, prescriptions, confidential letters, medical images, imaging reports, biological reports, other types of medical reports, etc. All the mentioned clinical documents include diverse health information, among which we distinguish sensitive information that is considered as highly Protected Health Information (PHI) [1]. Personal healthcare information is not only used in healthcare practices and shared between healthcare professionals, but also in public practices and research activities such as public health surveillance and public health research. Public health practices and research present risks that are related to the unauthorized disclosure of PHI [2]. Therefore, it is crucial for healthcare organizations to ensure PHI protection and to preserve the privacy of individuals. Particularly, the individual's privacy protection is required by legislation, such as the Health Insurance Portability and Accountability Act (HIPAA) legislation [3][4] and the GDPR (General Data Protection Regulation) [5] on personal data protection. Consequently, privacy requirements should be respected and ensured when designing systems and procedures for health data management. For that, data oriented care pathway models

present a highly required mean of sensitive data discovery within the shared clinical documents. Furthermore, using a standard structure of medical documents facilitates the discovery and the protection of the included PHI. Medical data classification using ontology for the clinical documents architecture could provide more data fluidity in order to implement personal data protection law. This facilitates the map of each level of protection to a set of privacy requirements as demanded by international health data legislation namely, the Health Insurance and Accountability Act (HIPAA). We believe our model could enhance compliance with privacy with regards to the protection of sensitive patient data within hospitals. Furthermore, the model we are presenting in this paper, describes the different clinical workflows typically included in hospitals care pathways as well as the set of input and out-put data for each workflow, which simplifies the sensitive data discovery task.

Our approach suggests a set of steps towards a legally shared HL7 clinical documents architecture ontology based on the privacy by design principle, which means the implementation of privacy requirements since an early stage of healthcare information systems design with respect to carrying out clinical pathways. In this paper, we take the osteosarcoma clinical pathway as a case study to validate our approach. The details of our approach are as follows:

- Model medical care pathways as business processes that emphasise shared clinical data aiming to identify sensitive health information among them.
- Identify the privacy requirements and procedures for each type of sensitive health data identified within the business process representing each care pathway.
- Identify a clinical data model based on the business process modelling.
- Define the sensitive health information categories.
- Define the HIPAA legislation requirements to preserve the patient's privacy and confidentiality with regards to the use of their PHI.
- Model the clinical pathways based on business process modelling in order to extract the shared clinical documents between healthcare professionals.
- Map clinical documents' data to a set of attributes as a step towards a clinical data ontology development.
- Identify the PHI underlining each process model.
- Define privacy requirements for PHI protection from any disclosure or misuse.

This paper is divided into sections as follows: in Section II, we present the related work; in Section III, we present the clinical pathway subject to study, as well as the clinical pathway modelling language of our choice. We adopt a data driven business process clinical pathway modelling approach. In Section IV, we present our clinical document architecture. In Sections V and VI, we define the privacy requirements for PHI, followed by results and discussion. In Section VII, we present the conclusion and future work.

II. RELATED WORK

In the literature, the study of clinical care pathways is vast due to the diversity of the modelling approaches, the analyses purposes and the care pathway level of specification. Depending on the concerned field, four groups of modelling techniques are defined [6][7][8]: 1) *Statistical and mathematical techniques*, based on the finding of significant relations among variables in order to assess the relation between patient's identifiers and their medical history [8], 2) *Data mining for clinical pathways*, based on discovering patterns in the medical events sequencing in order to predict the outcome of the next step of the care pathway [7], 3) *Business process modelling*, the modelling stage relies on a domain specific ontology for presenting systems in order to be analyzed and improved, it gives attention to the medical resources and documentation presentation in addition to the control-flow [9] [10] and 4) *Process mining algorithms*, dedicated to the processes analyses [11].

The modelling approach is based on questionnaires collecting information throughout interviews with doctors. They do not consider any data source and are subjective. For that, clinical pathways are textually and medically described. Their business processes and workflows are mostly detailed by doctors using textual description of the sequenced tasks. Due to the technological revolution in the medical field that includes medical information systems, several methods and business process modelling languages were suggested. This includes the Integration DEFinition language (IDEF) (V.0 and V.3), the Unified Modelling Language (UML) V.2.0 and the Business Process Model and Notation language (BPMN). Most of these technologies were also used to model clinical pathways [12].

BPMN is the most widely used and accepted language in medical process modelling thanks to its simple and high-level process construction. Clinical pathways processes are known as complex to be understood by medical practitioners. This needs a transparency of the whole process elements such as structures, participants, tasks, roles, etc. Modelling care pathways in the form of clinical processes is considered as a solution to overcome its complexity and define its requirements with regards to patients and health care professionals' needs. Therefore, medical process models should be simple, transparent and understandable as much as possible [13][14].

Despite the importance of business process modelling in clinical pathways and efforts for processes' automation, few works are dealing with care pathways' automation. Most of them are relying on a business process-based modelling approach. Besides, there is some suggested BPMN extension implementation such as the Clinical Pathway (CP) extension of the BPMN, called BPMN4CP, which proposes an ontology based- extension for e-health process

management. Other existing research works offer clinical textual description of the care pathways processes. In addition, other works were interested in analyzing systems behavior throughout business process-based modelling using UML, particularly, UML class diagram [15][16][17].

However, less effort has been made in investigating approaches for clinical data modelling with special interest in privacy preservation. In this context, our work is addressed to the respect of privacy requirements since a very early stage of HIS design in order to ensure a protected rolling of data driven business processes that are clinical pathway-oriented. Based on the modelling approach and the implementation of the *Business Process Modelling* technique, we highlight shared medical data throughout document centric clinical care pathway modelling and HL7-CDA standard investigation [18]. This is for presenting a shared data model and mapping them to a set of characteristics aiming to develop an ontology representing data related to the care pathway field. This will facilitate the definition of the required security level for each data type with respect to personal data protection legislation.

III. CLINICAL PATHWAYS

Clinical pathways are acknowledged as complex processes due to the diversity of the participating entities (e.g., healthcare professionals and medical service providers). Throughout the literature exploration of the clinical pathways' modelling and automation, we extracted the main phases underlining care pathway processes. A generic clinical pathway begins by an admission phase in which the patient is allowed to get access to the care establishments. This is usually followed by a diagnosis phase: that describes the visiting of the consulting doctor and having clinical diagnosis performed. The treatment phase should then occur: after identifying the pathology in the second phase, the treating doctor identifies the treatment protocol. This clinical pathway ends with the follow-up phase which allows the involved practitioner to monitor and evaluate the effectiveness of the prescribed treatment or to control the pathology progression [19].

The clinical pathway is a set of processes and sub-processes in which one or more healthcare professionals participate. The business process modelling of clinical pathways allows to identify the tasks, the participants and their roles in the care pathway proceeding. Even the shared data between healthcare professionals may also be modelled and identified among a clinical pathway-oriented business process [20].

In the following sections, we will detail the clinical care pathway of osteosarcoma and describe a step by step methodology to model our clinical business process and sensitive data discovery.

A. An Overview of Osteosarcoma Clinical Pathways

Osteosarcoma is a bone cancer. It most commonly reaches those aged from 10 to 30. A great part of this affected population concerns teenagers. Each year, from 800 to 900 people are estimated to be diagnosed with osteosarcoma in the United States. Osteosarcomas are primary malignant bone tumors. They can be classified according to cells' behavior under the microscope as high, intermediate or low grade [21].

Osteosarcoma clinical care pathways are characterized by their complex and multidisciplinary procedures with their difficult management facts. By Ferrante [13], the osteosarcoma first diagnosis starts with symptoms appearances like bone pain or soreness, a felt mass through the skin, swelling and redness, etc. During the clinical pathway diagnosis phase, while an osteosarcoma is suspected some standard imaging exams must be performed. Once the osteosarcoma is confirmed and its malignancy is not excluded, a biopsy should be performed allowing the cancer staging. As a final checkup step, several imaging exams are performed to verify the existence of metastases. A percentage of 85% indicates that the most common metastases appear in the lung whereas the bone is considered as the second most common site of distant disease [22].

The osteosarcoma checkup and grading steps allow the choice of the treatment procedure which includes chemotherapy, radiation therapy and surgery operation. The identification of the right therapy protocol is based on biological analyses. To verify and evaluate the treatment efficiency, the patient has to be periodically followed-up. This osteosarcoma clinical pathway follow-up step is based on the performing of imaging exams in addition to biological analyses as needed [22][23][24].

The complexity of the osteosarcoma clinical pathway business process is due to the collaboration between healthcare professionals from several medical departments. This process cannot be accomplished without clinical data sharing and transmission. Among those data, we find various sensitive data which are individually identifiable that are transmitted or maintained in electronic media [25]. For that, it is necessary to respect the applicable data protection regulation. For that, we need data driven clinical pathways' models, which facilitate the data discovery step and mapping them to sensitive data protection principals. Those models allow defining shared and transmitted clinical documents throughout medical processes. The clinical care pathways modelling step is based on the clinical processes description within the literature while the data discovery step is based on the investigation of the HL7-CDA standard. Then, we defined sensitive data according to the HIPAA regulation. As a consequence, the sensitive data discovery step is fulfilled according to both HL7-CDA standard and HIPAA regulation.

B. BPMN As Clinical Pathway Modelling Language

To model clinical pathways, we used the BPMN as a modelling language. It is the most widely used language in healthcare business process modelling. First, we explored

the clinical healthcare pathways in the literature on description of clinical pathways. Then, we divided them into three main phases, diagnosis or check-up, treatment and follow-up, mentioned above. Throughout the studied clinical pathways, we present the osteosarcoma clinical pathway as a case study to illustrate our data driven clinical healthcare pathway model. In order to elaborate the clinical pathway data driven model, we used the common patterns and symbols of the BPMN modelling language [12].

C. Osteosarcoma Clinical Pathway Modelling Using BPMN

In order to identify the clinical data that may be transmitted and shared between healthcare professionals as required by standard care pathway specifications, we modelled the osteosarcoma clinical pathway in the form of a business process model. In this way, we could first identify the characteristics of performed clinical tasks. Then, we could highlight the data driven tasks for the chosen pathology. The sections below present the data driven clinical pathways of osteosarcoma for the check-up, the treatment and the follow-up phases respectively.

1) Osteosarcoma checkup clinical pathway

Fig. 1 presents the check-up phase of the osteosarcoma clinical pathway as well as the shared and transmitted clinical documents, ensuring the steps required by the care pathway definition. The osteosarcoma clinical pathway is complex and involves collaboration of healthcare providers and diverse medical services including radiology, biology, nuclear medicine, surgical units, etc., as shown in Fig. 1, Fig. 2 and Fig. 3. After the patient admission, a medical consultation takes place. According to the clinical examination, medical tests are performed to accomplish the diagnoses phase and precise the pathology Fig. 1. The fulfilment of this phase needs the transmission and the share of diverse medical documents types (e.g., reports, orders, images, etc.) and subtypes (e.g., imaging report, anatomic pathology report, etc.) between healthcare providers. According to the HL7-CDA standard, each clinical document has a predefined standard structure [18].

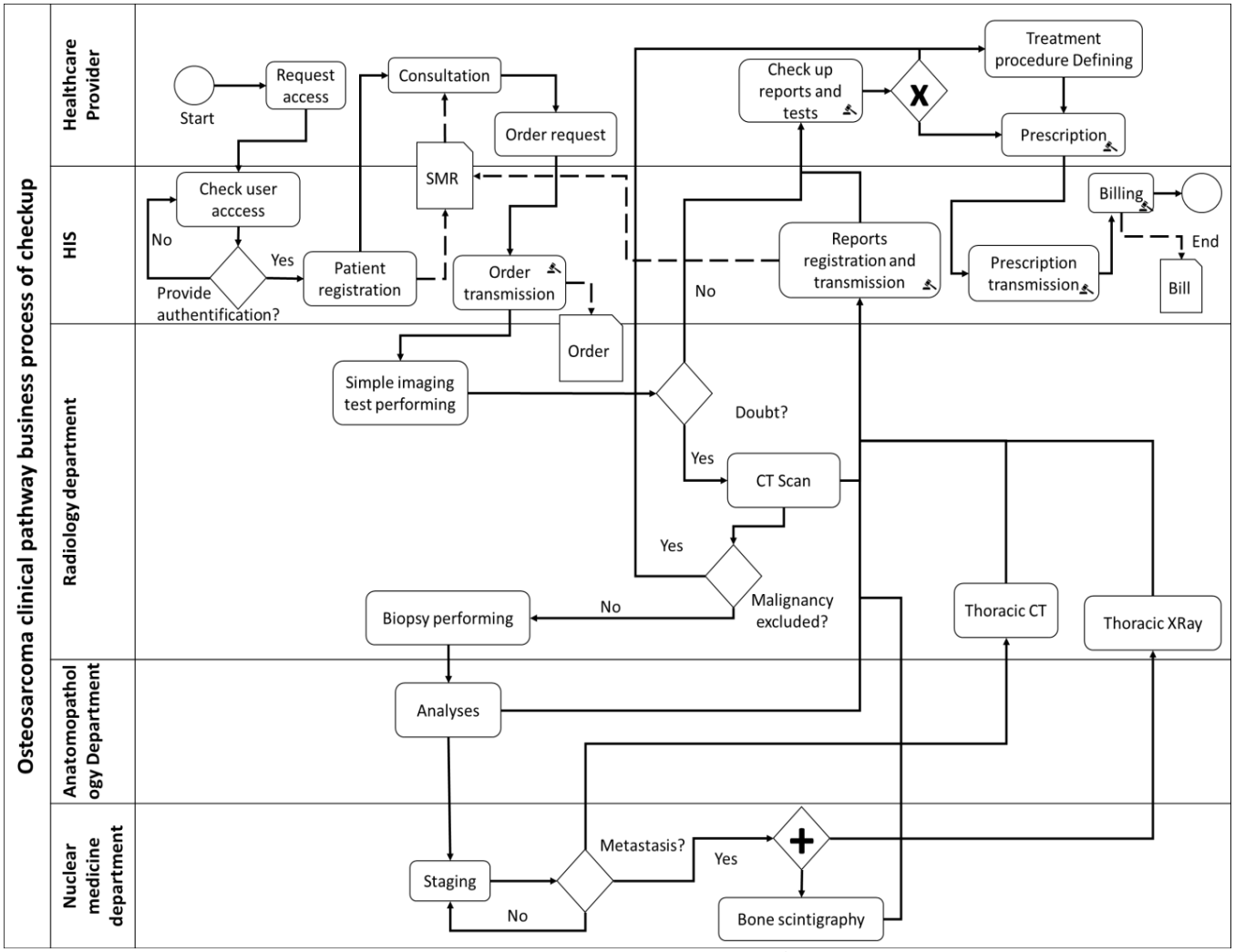

Figure 1. Osteosarcoma clinical pathway business process of checkup

2) Osteosarcoma treatment clinical pathway:

Based on tests' findings within the diagnosis phase clinical documents, the doctor defines the treatment phase according to systems review by biological tests, audiogram hearing tests and heart tests. By Luetke [23], during the treatment, medical tests and clinical examination are performed to evaluate its effectiveness as shown in Fig. 2. As the previous phase of the osteosarcoma clinical pathway, the treatment phase needs that many healthcare providers to get involved in it in order to be successfully fulfilled. Shared and transmitted medical documents in the course of this phase processes may be used for many purposes as the public healthcare and the public research. Besides, those documents are generally transmitted to the cancer registry in order to elaborate statistics. For that, the impact of the findings in medical documents (e.g., findings in chemotherapy, surgical operations and radiotherapy) on care quality improvement is very important. As a consequence, data discovery presents a crucial step to extract useful data for the medical data secondary use. This allows the sensitive data discovery which need to be protected in compliance with the HIPAA regulation.

3) Osteosarcoma follow up clinical pathway:

The last phase in the osteosarcoma clinical pathway is the following-up phase presented in Fig. 3. According to Paiolil [24], the doctor follows the patient health status by performing some tests and medical examination to check periodically the treatment effectiveness. Findings in medical documents managed in this phase may be used also for secondary use. Throughout the three phases of the osteosarcoma clinical pathway, diverse clinical documents are shared, transmitted and updated within the Shared Medical Record (SMR) ensuring the healthcare continuity.

Furthermore, clinical documents management enforced by Hospital Information Systems increases the risk of sensitive data disclosure or misuse. Despite the enhanced security measures while implementing network infrastructure and Hospital Information Systems, data hackers still find ways to penetrate networking systems and hack data. As a consequence, for a secondary use, the risk of sensitive data disclosure or misuse increases too. For that, we have resorted to sensitive data discovery managed throughout clinical pathways. This allows reinforcing its protection in compliance with the HIPAA regulation by predicting the risk rate of data disclosure that each type of medical document faces, then implementing computerized security methods in Hospital Information Systems since an early stage of their design for ensuring privacy enhancement.

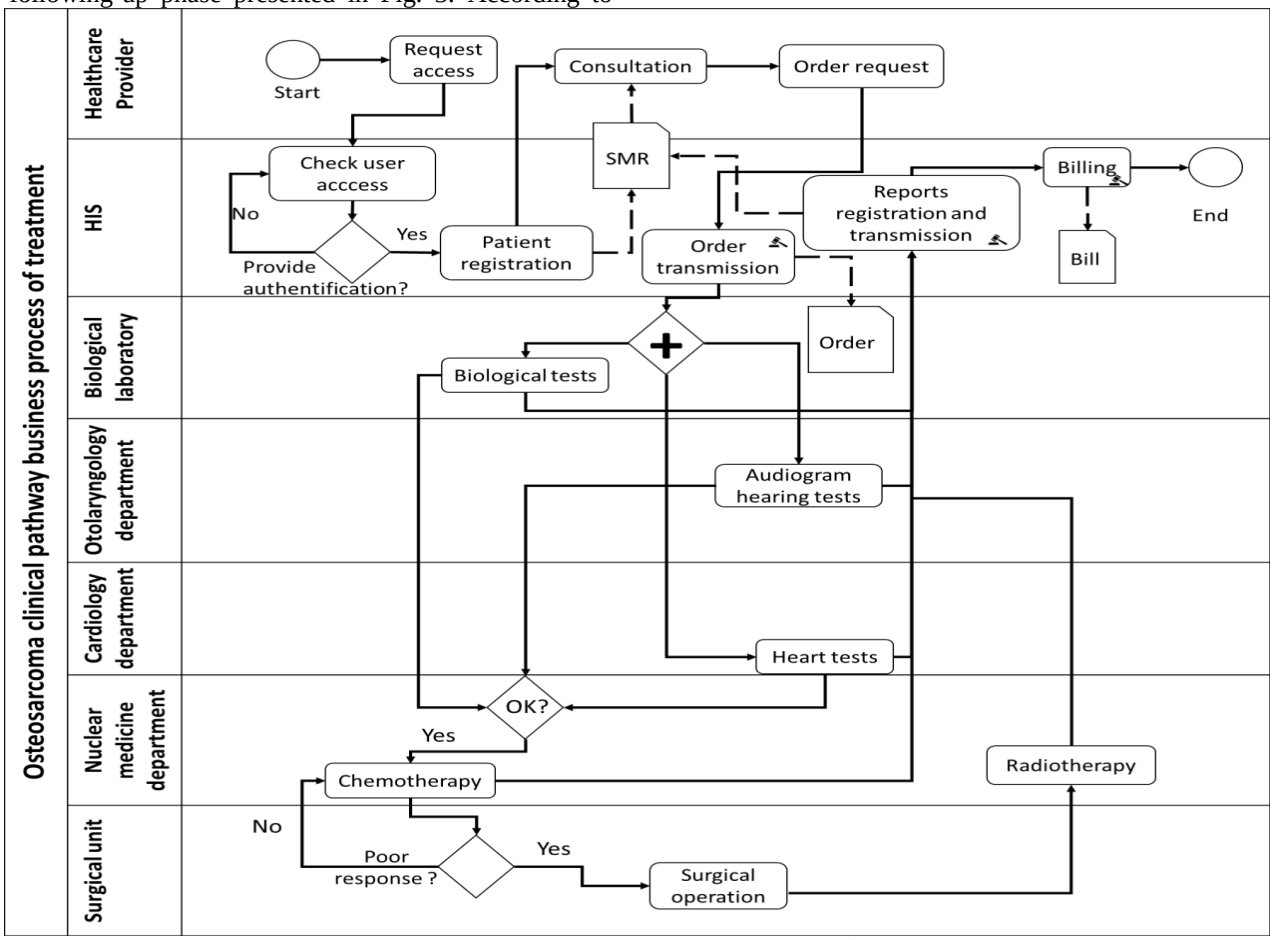

Figure 2. Osteosarcoma clinical pathway business process of treatment

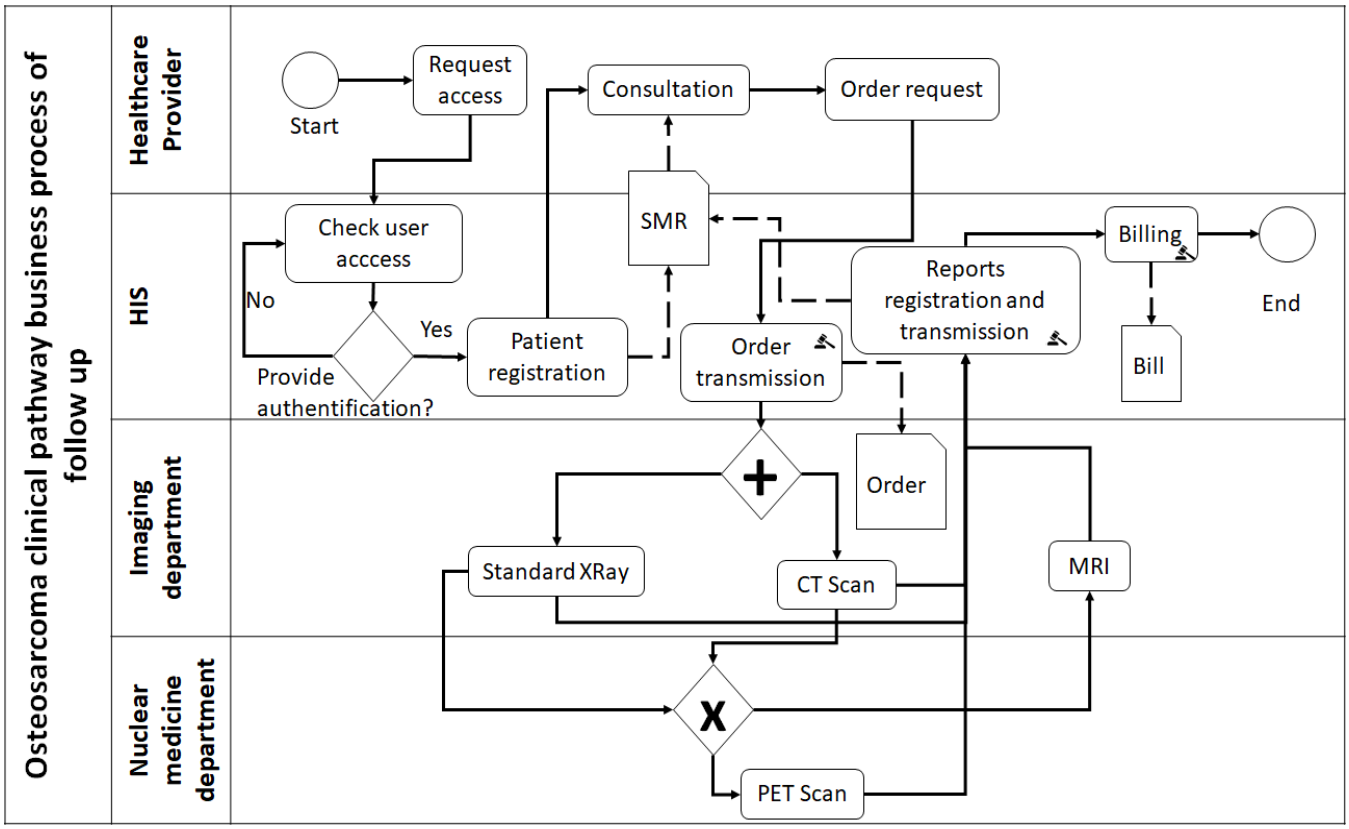

Figure 3. Osteosarcoma clinical pathway business process of follow up

IV. SHARED CLINICAL DOCUMENT ARCHITECTURE

A patient's Electronic Health Record (EHR) must contain all types of clinical documents including their medical history record, discharge summaries, typical paper charts, mental status examinations and other medical reports such as medical tests and operative reports.

Throughout a clinical business process, the EHR is transferred, updated and shared between healthcare professionals ensuring the continuity of care. Each clinical document included in the EHR contains medical data as it is required in the concerned healthcare establishment.

The general clinical document architecture is divided into documents, fragments and data. As shown in Fig. 4, clinical documents are composed of many fragments. They provide information about patients, procedures, practitioners, diagnosis, findings and appointments. The clinical shared documents' architecture model, illustrated in Fig. 4, could be adapted to another health care establishment, according to the used medical documents' structure in their boundaries. In each clinical document fragment, several medical data are found with specific properties which need the implementation of privacy by design approach. This is dedicated to the PHI use and disclosure within the HIS. The identification and demographic fragments in clinical documents include PHI. Its use should obey to the data protection law principles and privacy requirements ensuring the PHI privacy and the security of the medical data in use [26][27].

Based on data discovery using the HL7-CDA standard, the identification of sensitive data among the shared and the transmitted clinical documents facilitates the application of the required security measures. This could be applied by the use of security computerized methods (e.g., encryption, decryption, anonymization and pseudonymization) since an early stage of the design of the HIS.

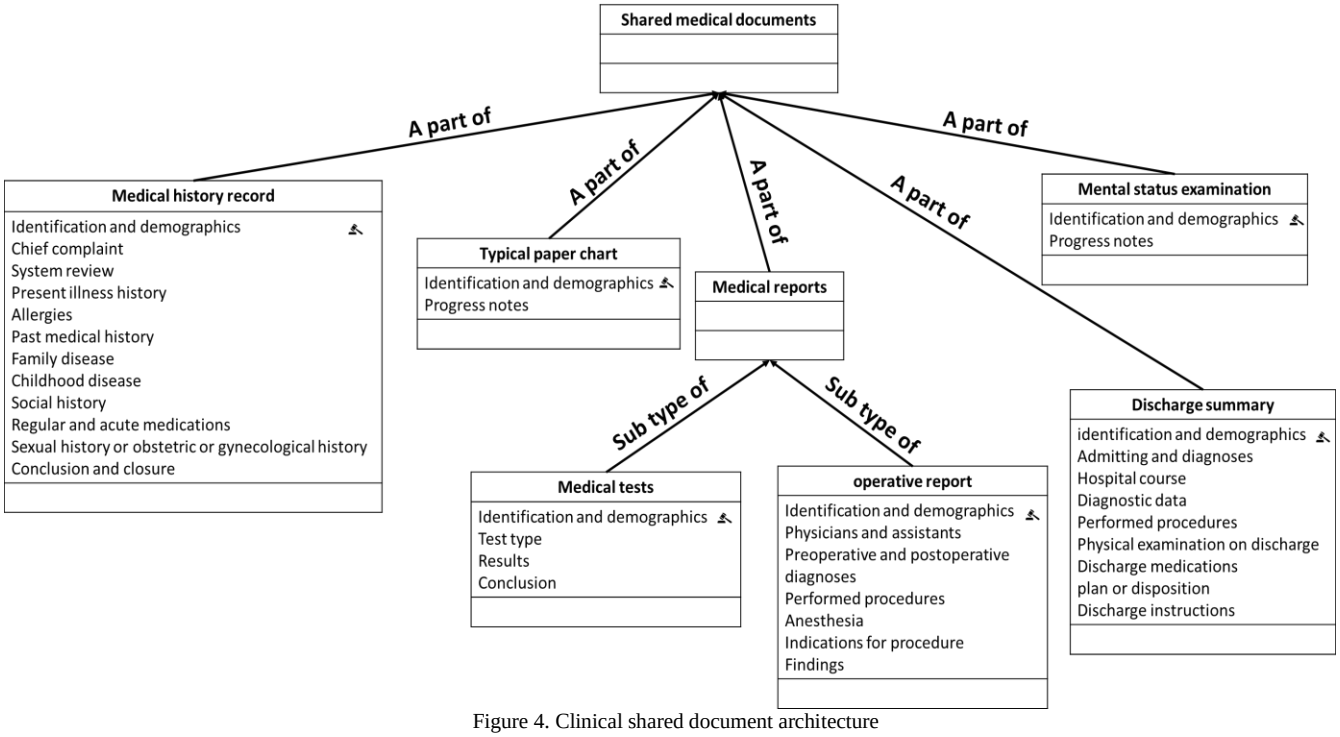

Figure 4. Clinical shared document architecture

V. HL7 CLINICAL DOCUMENTS ARCHITECTURE MAPPING

Data driven business process modelling enabled us to highlight shared clinical documents throughout different clinical workflows typically included in hospitals care pathways. Referring to the HL7-CDA standard documentation [18], we were able to investigate and define the shared clinical documents architecture as well as the included attributes. Thanks to the personal data protection principals, the PHI privacy requirements included in the HIPAA rules and the PHI defined within the HIPAA, we were able to discover sensitive data and highlight the PHI included in the shared clinical document.

The use of standardized architecture of clinical documents could facilitate distinguishing PHI from other included data. Furthermore, it facilitates the implementation of data protection methods while insuring a simplified clinical documents management and processing. The ontological representation of clinical documents' architecture allows capturing clinical information in the form of medical terminologies. Besides, it allows the classification of the clinical data into categories and the mapping each data category to a set of adequate data protection mechanisms as required by HIPAA Privacy Rule [3][4]. The use of ontologies helps us to deal with the nature of realism; it is the science or the study of being [28]. It allows the description of the included data in the clinical documents regarding to the human logic and philosophy to analyze the reality of things [29]. It generally provides a domain or a subdomain knowledge representation. A data model representation using ontologies generally provides a formal and conceptual model describing the logic of the data structure as well as the meaning of the data elements, which allow it to be understandable to both human and machine. Ontology-based models usually describe individuals (instances or objects), classes (concepts or types of objects), properties (attributes) and relations in a particular domain. Those relations describe ways in which classes and individuals could be related to one another [28].

Based on the CDA (Clinical Document Architecture) documentation of the HL7 standard, we mapped clinical documents to a set of characteristics aiming to provide a mean of classifying the included data in shared clinical documents. Each clinical document is characterized by a set of attributes providing information about the clinical document name, the category of each document (e.g., medical reports, medical record, etc.) and the clinical document architecture. This allows sensitive data discovery within shared clinical documents. After deep study of the general architecture of a clinical shared document, we could classify the included content into a set of metadata, data and values as shown in Table I and in Fig. 5.

The clinical document architecture is divided into two main parts: a header and a body. The first part of the document provides information about the document title, the document type, the document version, the participants, the organization, etc. As for the second part, it provides information about the clinical document content as the findings, reason for study, history, procedure context, study act and observation. So, we could map the clinical documents' data to attributes, as described in Table I. In this

part, we are taking an imaging report as an example to illustrate the clinical documents architecture according to the HL7-CDA standard [30]. The chosen document belongs to the category of *medical reports*. Particularly, it is a part of the *medical tests* as represented in Fig. 4. Most of the data values are associated with some standard variables part of CDA. However, many other variables are user defined to allow adaptation of the process model to specific organisational contexts. Some data could have suggested possible values as for the case of the *clinical document* type which can take a *CT Report* as a value for example. The included data in a clinical document has a hierarchical structure that specifies the recommended information for each shared document, participant, authorization, clinical statements and each section required to be present in this document according to its type. For example, for the *participants* section in the header of the imaging report, we found *record target* representing the patient's information, *author*, *data enterer*, *information recipient*, *legal authenticator* and *participant* as sub sections. For each sub section, a set of data values are required to be affected while the document processed and managed within the patient care pathway building the medical processes of the hospital information system. As it is illustrated in Fig. 5, the *record target* (patient) sub section affords information about the *patient role* which includes *address*, *phone number*, *patient name*, *birthdate* and his *gender*.

Data mapping and classification based on HL7-CDA standard presents the first step of our sensitive data discovery approach. As for the second step, it is based on the investigation of personal data protection principals and the extraction of the defined PHI within the HIPAA Privacy Rule as described in the following section. Matching findings of the first step with those of the second one allowed the sensitive data discovery within data driven clinical pathways.

TABLE I. CLINICAL DOCUMENT ARCHITECTURE DATA MAPPING

Characteristics	Attributes	Metadata	Data	Values
Clinical Document Name	Name	_	Value of the document name	Imaging report
Category	Category to which belong the clinical document	_	Value of the document category	Reports
Clinical Document Architecture	Header	Clinical Document	Title, date (effective time), version	Values are affected within the hospital information system
			Type	- Diagnostic Imaging Report - CT Report - MRI Report - Ultrasound Report - Nuclear Medicine Report - PET Scan Report - Cardiac Catheterization Report - Echocardiography Report - Colonoscopy Report - Endoscopy Report - Electrophysiology Report - Obstetrical Ultrasound Report
		Participants	Record target (patient), author, data enterer, information recipient, legal authenticator, participant	Values are affected within the hospital information system
		In fulfillment of	Order	
		Documentation of	Service event, physician reading study performer	
		Authorization	Information about the authorization	
		Related document	Parent document	
		Component of	Encompassing encounter	
	Body	Dicom Object Catalog	Study, series, SOP Instance UID	Values are affected within the hospital information system
		Findings	Sections (include paragraphs)	
		Optional sections	Reason for study, history, impression	
		Clinical statements	Procedure context, study act, text observation	

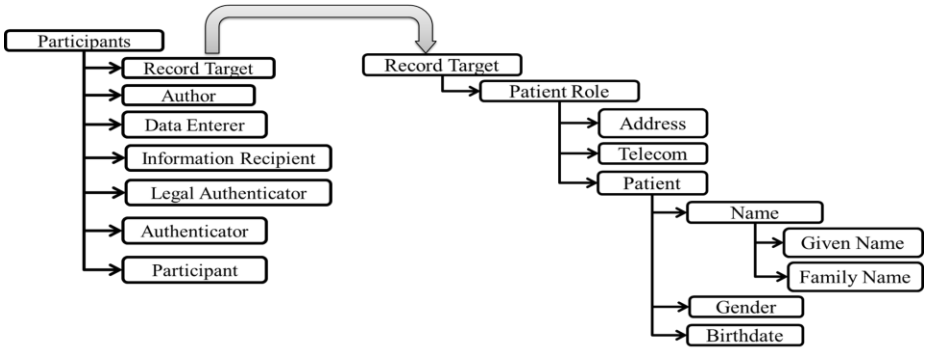

Figure 5. Example of the hierarchical structure of the participants section within an imaging report

VI. PRIVACY REQUIREMENTS FOR PHI PROTECTION

The use of clinical healthcare data is governed by many jurisdictions as it may present risks threatening a person's life and may affect both his/her privacy as well as his/her professional life. For this, clinical data usage must be set for data protection principles. In particular, the following eight principles should be respected as required in HIPAA regulation and the European directive:

1- Lawfulness, fairness and transparency: personal data should be processed lawfully, fairly in a transparent way.
2- Purpose limitation: personal data should be processed for specific purposes.
3- Data minimization: personal data should be adequate, relevant and limited to the precise purposes.
4- Accuracy: personal data should be kept up to date.
5- Storage limitation: personal data should be kept for no longer than the necessary period for the purposes for which those data are processed.
6- Rights: people have the right to access their data and give permission for other entities to use or disclose them.
7- Integrity and confidentiality: personal data should be processed in a secure way. They should be protected also against any unauthorized or unlawful processing, accidental loss, destruction or damage.
8- International transfers: personal data should not be transferred outside countries [26].

International law frameworks, such as European directive and HIPPA for personal data protection are based on the previous data protection principles. The present work is developed with regard to Protected Health Information within HIPAA regulation. The HIPAA Privacy Rule is published by the department of Health and Human Services (HHS) to ensure health information privacy. The privacy rule is applied to covered entities as health plans, healthcare clearinghouses and the healthcare providers. It defines a set of rules in order to protect sensitive health information with respect to its use and disclosure. Sensitive health information is known as Protected Health Information (PHI). They are individually identifiable health information related to the patient's past, present and future physical or mental health conditions, the healthcare provision to the individuals and the past, present or future healthcare provision to individuals [2]. The individually identifiable health information includes demographic data and many common identifiers. PHI usage and disclosure are permitted without the patient's informed consent for some purposes and situations as to the individual, the treatment, payment and healthcare operations, opportunity to agree or object, incidence to an otherwise permitted use and disclosure or public interest and benefit activities as well as a limited data set for research, public health or healthcare operations purposes or when it is required by law. As for the not permitted PHI usage and disclosures, an individual's written authorization (consent) must be obtained [3].

In addition to permitted PHI use and disclosure, prohibited ones are defined in Privacy Rules. For example, genetic information is considered as PHI and they shall not be used or disclosed for underwriting purposes as well as the psychotherapy notes. Furthermore, PHI may not be sold by covered entities. The PHI use and disclosure must be limited to the minimum necessary. However, PHI may be used to create a non-individually identifiable health information or a de-identified information [3][4].

The HIPAA Privacy Rule also defines a set of PHI de-identification requirements in order to use and disclose it without the patient's authorization. A covered entity may de-identify PHI by removing the eighteen identifiers specified in the following list as defined in HIPAA Privacy Rule:

1. Names.
2. Addresses with all geographic subdivisions smaller than a State.
3. Dates except year (birthdate, admission and discharge date, date of death).
4. Telephone numbers.
5. Fax numbers.
6. Email addresses.
7. Social security numbers.
8. Medical record numbers.
9. Health plan beneficiary numbers.

10. Account numbers.
11. Certificate/license numbers.
12. Vehicle identifiers, serial numbers and license plate numbers.
13. Device identifiers and serial numbers.
14. URLs (Web Universal Resource Locators.
15. IP (Internet Protocol) address numbers.
16. Biometric identifiers (finger and voice prints).
17. Full face photographic images and any comparable images.
18. Any unique identifying number characteristic or code [1].

For the above identified PHI usage and disclosure purposes, de-identification based on computerized methods is necessary to respect the PHI privacy and ensure its protection from any illegal use or other threatening risks. Furthermore, they allowed us to identify the PHI included in the shared clinical document with regard to a set of clinical document data attributes as described in Section V. Our sensitive data discovery is based on the defined PHI in HIPAA Privacy Rule. They could be defined as sensitive information while our clinical document architecture ontology development. This will facilitate clinical documents and data mapping to a recommended data protection techniques that ought to be HIPAA compliant. For that, we will first start by defining generic data protection techniques as required by HIPAA then we will highlight the applicable techniques according to the clinical document type and the contained sensitive data.

VII. RESULTS AND DISCUSSION

In this present work, we are interested in discovering sensitive data within the shared and transmitted medical document throughout clinical business processes. Therefore, we studied medical business processes in order to elaborate a data driven clinical pathway model, based on the BPMN language. Then, we mapped the clinical documents included data to a set of characteristics in order to provide a meaningful data classification. Based on the HIPAA Privacy Rule, we extracted personal data protection principals as well as the eighteen identifiers defined as Protected Health Information. The aim here is to ensure the respect of privacy requirements since early stages of HIS design. Then, we divided clinical data into categories and extracted PHI among them in the form of data model clinical pathway. After that, we defined both personal data protection principles and HIPAA privacy requirements for the specified PHI use and disclosure. Finally, all of the above listed objectives were validated through the modelling of osteosarcoma care pathway business process model chosen as a case study, mapping clinical data to a set of characteristics and discover sensitive data among them in order to facilitate our clinical data ontology development.

As for the completion of the modelling phase of osteosarcoma clinical pathway, we have modeled its complex care pathway which is divided into three phases: *check-up*, *treatment* and *follow-up*. This was done using the actual BPMN language simple patterns. Hence, personal data processing is integrated in the processes, particularly, in a legislation compliant manner which adds more trust to medical documents processing and sharing during the clinical process implementation. The shared clinical document data mapping to sets of attributes has allowed discovering sensitive data and classifying clinical data into formal concepts that are clearly structured and outlined. Data protection principals' definition according to the HIPAA Privacy Rule allowed the identification of sensitive clinical data or PHI included in shared clinical documents. This data ought to be protected using HIPAA compliant data protection techniques within the clinical document processing and management throughout clinical business processes and workflows. Data protection techniques will be defined according to the shared clinical documents' type and structure within the HL7-CDA standard and the HIPAA Privacy Rule compliance.

Many difficulties were encountered in clinical pathway modelling using BPMN due to the complexity and multidisciplinary aspect of medical procedures. This has led us to conclude the necessity of a more specialized care pathway modelling language. This has also highlighted the need for a new care pathway modelling and automation language that is sensitive-data driven and could integrate privacy requirements specification. Thus, a new extension of the BPMN modelling language is required.

VIII. CONCLUSION AND FUTURE WORK

Clinical pathways automation is highly required in standardized HIS. This is traditionally ensured by business process modelling. In this context, we developed a data driven clinical pathway business process model for osteosarcoma, as a case study. We used BPMN as clinical pathway business process modelling language. A shared clinical data model was elaborated further to the clinical business process model. Based on the HL7-CDA standard, we were able to discover clinical data and define its structure. Moreover, the mapping of the obtained data to a set of clinical document elements was necessary to specify the logic of the shared clinical data formal representation. This facilitates sensitive data discovery and PHI highlighting within the shared clinical documents referring to PHI defined in HIPAA Privacy Rule. This allows the application of personal data protection techniques in a more fluid compliance to HIPAA.

Since personal data management must obey to data protection law, we defined both personal data protection principles and HIPAA privacy requirements with relation to patients identifying medical documents, in terms of their both use and disclosure.

The adoption of privacy by design approach offers a better enforcement of privacy since an early stage of computer-based healthcare systems design. This allows an orthogonal integration of privacy obligations throughout the

clinical process. For this reason, we are working currently on sensitive data discovery within shared and transmitted clinical documents between healthcare providers. Mapping discovered data to a set of personal data protection requirements. Then, we identified the defined PHI within HIPAA Privacy Rule. This was very useful to discover sensitive data that need more protection than other shared data. This is crucial step to achieve our clinical documents' ontology development and apply HIPAA compliant measures. BPMN process modelling language need to be extended with privacy annotation features and additional patterns to allow modelling privacy specification as part of clinical processes and giving them more attention since an early stage of the HIS design. The definition of a common vocabulary qualifying clinical pathways specifications will more enforce the respect of privacy requirements. We believe clinical process modelling languages should be more adapted to a multidisciplinary clinical systems users' profile. The adoption of a variety of symbols and modelling patterns investigation is needed in order to better tailor to the requirements of the clinical community.

Sensitive data discovery based on data driven clinical business process models and PHI protection requirements within HIPAA compliance facilitated distinguishing sensitive data from other medical documents included data. This provides a mean of classifying the included data in shared clinical documents. Thus, it facilitates our clinical data ontology development which provides a simple way to associate the security level and risk to each sensitive data category.

REFERENCES

[1] I. Essefi, H. Boussi Rahmouni, and M. F. Ladeb, "Data Driven Medical Process Modelling for Privacy Protection in Care Pathways," The Seventh International Conference on Global Health Challenges (GLOBAL HEALTH 2018) IARIA, Nov. 2018, pp. 24-31, ISSN: 2308-4553, ISBN: 978-1-61208-682-8.

[2] U.S. Department of Health & Human Services *HHS: HIPAA Privacy Rule Summary*. [Retrieved: July, 2013]. http://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[3] HIPAA Privacy Rule: Uses and disclosures of protected health information: General rules (§164.502). [Retrieved: January, 2013]. https://www.law.cornell.edu/cfr/text/45/164.502.

[4] HIPAA Privacy Rule: Uses and disclosures for which an authorization is required (§164.508). [Retrieved: January, 2013]. https://www.law.cornell.edu/cfr/text/45/164.508.

[5] Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). [Retrieved: March, 2017]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0001.01.ENG&toc=OJ%3AL%3A2016%3A119%3ATOC.

[6] L. Bouarfa and J. Dankelman, "Workflow mining and outlier detection from clinical activity logs," Journal of Biomedical Informatics, vol.45, pp.1185–1190, 2012, Dec. 2012.

[7] Z. Huang, X. Lu, H. Duan, and W. Fan, "Summarizing clinical pathways from event logs," Journal of Biomedical Informatics, vol.46, pp. 111 – 127, Feb. 2013.

[8] S. Adeyemi, E. Demir, and T. Chaussalet. "Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes," Decision Support Systems, vol.55, pp. 117 – 125, Apr. 2013.

[9] R. Braun, M. Burwitz, H. Schlieter, and M. Benedict, "Clinical processes from various angles - amplifying bpmn for integrated hospital management," International Conference on Bioinformatics and Biomedicine (BIBM) IEEE, Nov 2015, pp. 837–845.

[10] M. Shitkova, V. Taratukhin, and J. Becker, "Towards a methodology and a tool for modeling clinical pathways," The 5th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2015) Procedia Computer Science 63, Sept. 2015, pp. 205 – 212, ISSN 1877-0509.

[11] A. A. Funkner, A. N. Yakovlev, and S V Kovalchuk, "Data-driven modeling of clinical pathways using electronic health records," Procedia computer science, vol. 121, pp. 835-842, Nov. 2017.

[12] E. Rolón, E. R. Aguilar, F. Garcia, F. Ruiz, M. Piattini, and L. Calahorra, "Process modeling of the health sector using BPMN: A case study," The First International Conference on Health Informatics-Diagnostic Pathology (HEALTHINF 2008) IARIA, Jan. 2008, pp. 173-178, ISBN: 978-989-8111-16-6.

[13] S. Ferrante, S. Bonacina, G. Pozzi, F. Pinciroli, and S. Marceglia, "A design methodology for medical processes," Applied Clinical Informatics, vol. 7, pp. 191-210, 2016, doi:10.4338/ACI-2015-08-RA-0111.

[14] F. Ruiz, F. Garcia, L. Calahorra, C. Llorente, L. Gonçalves, C. Daniel, and B. Blobel, "Business process modeling in healthcare," Stud Health Technol Inform, vol. 179, pp. 75-87, 2012.

[15] R. Braun, H. Schlieter, M Burwitz and W. Esswein, "Extending a Business Process Modeling Language for Domain-Specific Adaptation in Healthcare," The 12th International Conference on Wirtschaftsinformatik in Osnabrück-Smart Enterprise Engineering (WI 2015), Mar. 2015, pp. 468-481, ISSN: 0937-6429, ISBN: 978-3-00-049184-9.

[16] S. Bielack, D. Carrle, P. G. Casali, and ESMO Guidelines Working Group, "Osteosarcoma: ESMO clinical recommendations for diagnosis, treatment and follow-up," Annals of Oncology, vol. 20, pp. iv137-iv139, May. 2009, doi: 10.1093/annonc/mdp154.

[17] V. Augusto and X. Xie, "A modeling and simulation framework for health care systems," IEEE Transactions on Systems, Man, and Cybernetics Systems, vol. 44, pp. 30–46, 2014.

[18] HL7 standard. [Retrieved: 2019]. https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7.

[19] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," Journal of biomedical informatics, vol. 61, 224-236, June 2016, doi: https://doi.org/10.1016/j.jbi.2016.04.007.

[20] N. Hashemian and S. S. R. Abidi, "Modeling clinical workflows using business process modeling notation. In Computer-Based Medical Systems," The 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2012), June 2012, pp. 1–4, ISNN: 1063-7125, ISBN: 978-1-4673-2051-1.

[21] K. D. Miller, R. L. Siegel, C. C. Lin, A. B. Mariotto, J. L Kramer, J. H. Rowland, K. D. Stein, R. Alteri, and A. Jemal, "Cancer treatment and survivorship statistic," CA: a cancer journal for clinicians, vol. 62, pp. 220-241, Jun. 2016, doi: 10.3322/caac.21149.

[22] M. S. Isakoff., S. S. Bielack, P. Meltzer, and R. Gorlick, "Osteosarcoma: current treatment and a collaborative pathway to success," Journal of clinical oncology, vol. 33, pp. 3029-3035, Sep. 2015, doi: 10.1200/JCO.2014.59.4895.

[23] A. Luetke, P. A. Meyers, I. Lewis, and H. Juergens, "Osteosarcoma treatment–where do we stand? A state of the art review," Cancer treatment reviews, vol. 40, pp. 523-532, May. 2014, doi: https://doi.org/10.1016/j.ctrv.2013.11.006.

[24] A. Paioli, M. Rocca, L. Cevolani, E. Rimondi, D. Vanel, E. Palmerini, M. Cesari, A. Longhi, A. M. Eraldo, E. Marchesi, P. Picci and S. Ferrari, "Osteosarcoma follow-up: chest X-ray or computed tomography?," Clinical sarcoma research, vol. 7, Feb. 2017, doi: 10.1186/s13569-017-0067-5.

[25] HIPAA Privacy Rule: [Online]. Available .from: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html, last reviewed: 2015.

[26] M. C. Oetzel and S. Spiekermann, "A systematic methodology for privacy impact assessments: a design science approach," Eur. J. Inf. Syst., vol. 23, pp. 126–150, March. 2014, doi: https://doi.org/10.1057/ejis.2013.18.

[27] S. Rajamani, E. S. Chen, Y. Wang, and G. B. Melton, "Extending the HL7/LOINC Document Ontology Settings of Care," AMIA Annual Symposium Proceedings, Nov. 2001, pp. 994-1001, doi: https://doi.org/10.1016/j.cmpb.2015.09.020.

[28] C. Barki, S. Labidi, H. Boussi Rahmouni. "Ontology-driven generation of radiation protection procedures," World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering, vol. 11, pp. 256-261, 2017.

[29] B. Smith, "Beyond concepts: ontology as reality representation," The third international conference on formal ontology in information systems (FOIS 2004) IOS Press, Nov. 2004, pp. 73-84.

[30] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo, "HL7 Clinical Document Architecture, Release 2", Journal of the American Medical Informatics Association, vol. 13, pp. 30–39, Jan. 2006.