

# Structural Application and CRS Model Design for Information Protection in BIG DATA Analysis

<sup>1</sup>Nithiya.K, <sup>1</sup>PG student

Department of Computer Science and Engineering,  
<sup>1</sup>PG student, Parisutham Institute of Technology and Science,  
 Thanjavur, Tamilnadu, India.

**Abstract**— Due to the recent development of IT technology, the capacity of data has surpassed the Zetta-byte, and improving the efficiency of business by increasing the predictive ability through an efficient analysis on these data has emerged as an issue of the current society. Even each general hospital is getting a large amount of information stored and managed. However, there are very few studies on the methods to utilize the clinical information efficiently. For the efficient analysis and utilization of BIG DATA stored in heterogeneous electronic medical record (EMR) data, a data-cleaning process to facilitate the utilization of data is needed. To this end, there is a need for the process of cleaning clinical information composed of text to make sure that the data can be processed mechanically. In this regard, in this paper, medical information, medication information, medical test results and allergy information were implemented up to Entry-level using CDA, the international medical standards, and Care Record Summary integrating these information was created to ensure interoperability and enable more efficient medical treatment. In addition, CRS model which is suitable for the situation of Korea was designed, and a system that enables utilization of clinical data was proposed.

**Keywords**— CDA, Care Record Summary, HL7, Big Data, EMR

## I.INTRODUCTION

With the recent development of IT technology, the capacity of data has surpassed the Zetta-byte, and improving the efficiency of business by increasing the predictive ability through an efficient analysis on these data has emerged as an issue of the current society. Even in each general hospital, large amount of information is being stored and managed. However, there are very few studies on the efficient use of the medical information. For the efficient analysis and utilization of BIG DATA stored in heterogeneous electronic medical record (EMR) data, a data-cleaning process to facilitate the utilization of data is needed. To this end, there is a need for the process of cleaning clinical information composed of text to make sure that the data can be processed mechanically.

For the use of the cleaned clinical data, an environment that can easily take advantage of the clinical information is needed. In addition, since the clinical information includes personal

information, the anonymization of personal information that removes patients' personal information from the information is required before the analysis using clinical information.

As a document for the exchange of clinical information provided in HL7 (Health Level Seven), a American National Standards Institute, CDA (Clinical Document Architecture) has been used.[1][2] The CDA is divided into Section level and Entry level, and the CDA document that implements up to Entry level can be processed mechanically, and it ensures the interoperability due to the use of standard codes.

In this paper, medical information, medication information, medical test results and allergy information were implemented up to Entry-level using CDA, the international medical standards, and Care Record Summary integrating these information was created to ensure interoperability and enable more efficient medical treatment. In addition, CRS model which is suitable for the situation of Korea was designed, and a system that enables utilization of clinical data was proposed.

## II. METHODS

### Definition of CRS items and its design

The items were defined by analyzing the clinical information exchange items of top ranked hospitals in Korea and overseas CRS items. The clinical information exchange items of domestic hospitals (Samsung Medical Center, Asan Medical Center, Catholic University of Korea Seoul St. Mary's Hospital, Chungnam National University Hospital, Ajou University Medical Center, Hallym University Medical Center, Seoul National University Bundang Hospital) were analyzed, along with those of overseas (Canada[3], Singapore[4], the UK[5], the US[6], Australia[7]) ones, and the items of CRS were defined into patient information, diagnosis and medication history(inspection name, medicine information), past history, and surgical history.

CRS was designed using the CDA. As standards devised for the exchange of clinical documents, HL7 CDA consists of Header and Body, and the Body is composed of Section and Entry. The Header includes information on the patients, hospitals and date of medical treatment, and data on the clinical information is entered in the Body. The Entry enters the values of clinical information within the section in the Body as codes. HL7 provides Consolidated CDA Implementation Guide [8], a collection of templates of CDA documents used in the medical field. In this paper, the CRS was designed using the Consolidated CDA.

### B. CRS service methods

In Korea, all medical information of patients is not managed by the physician in charge unlike the practices in the United States, which poses the problem in determining the subject of updating CRS. This paper attempts to use a method to manage CAD documents generated by the occurrence of medical practices separately, not summarizing them. (Fig 1)

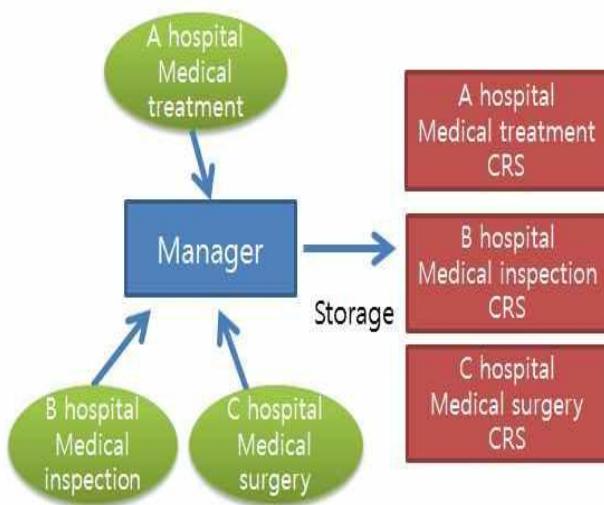


Fig. 1. CRS SERVICE METHODS

### C. Anonymization

Provision can be made after anonymizing identifiable information before the exchange of clinical information is done. The methods to anonymize are as follows.

- The value can be removed, or it can be substituted into separate value by checking attributes and elements that contain an identifier within the CAD document.

- The value can be removed, or it can be substituted into separate value by checking attributes and elements that include user identification information within the CDA document.

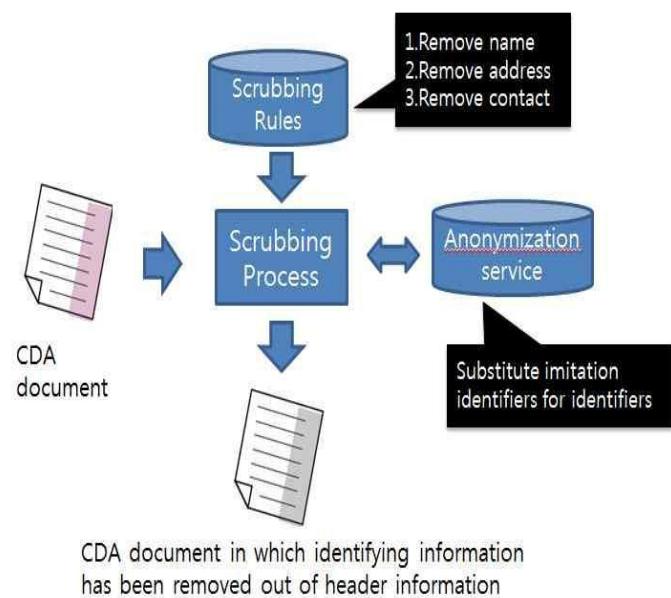


Fig. 2. Anonymization Method

## III. RESULTS

The use of LOINC(Logical Observation Identifiers Names and Codes)[9], SNOMED-CT(Systematized Nomenclature Of Medicine Clinical Terms)[10] standard, a universal code system by implementing up to CDA entry level enables interoperability between heterogeneous systems and mechanical processing of clinical information, which makes it easier to take advantage of the data. Table1 shows the location in which each item of CRS is mapped within the section of consolidated CDA.

TABLE I. CRS ITEM AND CONSOLIDATED CDA ITEMS

CRS Item	Consolidated CDA Item
Patient information	CDA Header
Medication information	Medication section
	-Medication Activity entry
	-Medication Information entry
	-Medication Supply order entry
Diagnosis	Problem section
	-Problem ConcernAct entry
	-Problem Observation entry
Inspection information	Result section
	-Results Organizer entry
	-Results Observation entry
Surgical information	Procedures section
	-Procedure Activity Act entry
	-Procedure Activity Observation entry
	-Procedure Activity Procedure entry
Vaccination history	Immunization section
	-Immunizations Activity entry
	-Immunizations Medication Information entry
Allergies and side effects	Allergies section
	-Allergy Problem Act entry
	-Allergy Observation entry
	-Reaction Observation entry
Visit history	Encounters section
	-Encounters Activities
	-ServiceDelivery Location

In order to make a good use of unstructured data stored in different EMR systems, the data needs to be stored in a structured form. To this end, this paper proposes a method to change the data to structured one using a CRS Data model and put it into a repository as shown Fig 4.

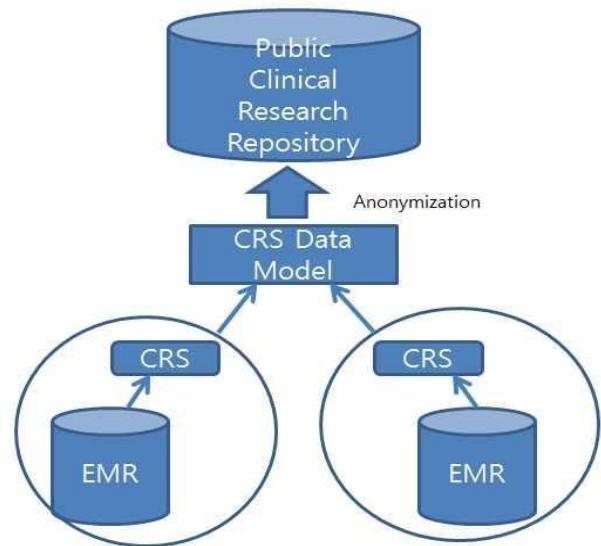


Fig. 3. CRS store process

CRSs are made from clinical data stored in the EMR system of each hospital. These CRSs are converted to fit into CRS Data Model and stored in the Repository. Since the information stored in the EMR includes personal information, it cannot be utilized as research data. Accordingly, it can be used after being stored in Clinical Research Repository through deleting the personal information of CRS in the process of conversion to fit into CRS Data Model.

#### IV. DISSUSSION

The CRS designed in this paper has its significance in that it proposed a method on how to structure the data using CRS as a method of converting unstructured medical information BIG DATA stacked on the each hospital into structured data that can be analyzed efficiently, thereby preparing a foundation for providing higher level of health care services by facilitating an analysis of the medical information BIG DATA and predicting the future.

Care Record Summary					
Patient Information					
name		birth date			
tel		occupation	Student	gender	Male
address					
Visit History					
Visit date	20111201	status	Type 2 diabetes		
Diagnosis name	diabetes				
Hospital name	seoul hospital				
medication information					
order date		drug name			
dose		time			
number of time					
diagnosis information					
date		Depressive disorder			
diagnosis name		diabetes			
status		Ongoing			
Allergies and side effect					
ATC above target					
Td due					

Fig. 5. The screen when inquiring about the anonymized CRS

The CRS also realized interpretability by enabling compatibility between heterogeneous systems through implementing up to Entry level using standard code of CDA, a clinical document makeup standard and enabled efficient medical treatment through exchange of medical information between hospitals(XDS environment)[11] by making it possible to view medical information at a glance. In addition, it is expected that it will help to promote efficient utilization of medical information and improve the quality of medical services by suggesting a system to store data after deleting personal information for the utilization of medical information and proposing a CRS service method that reflects the situation of Korea.

For future work, a study of methods to utilize and analyze the medical information shared, and a research on the terminology will have to be added. [12]

## REFERENCES

- [1] Dolin, R.H., Alschuler, L., Beebe, C., Biron, P.V., Boyer, S.L., Essin, D., Kimber, E., Lincoln, T., Mattison, J.E.: The HL7 clinical document architecture. *J. Am. Med. Inform. Assoc.* 8(6), 552–569 (2001)
- [2] Marcel Lucas Müller, Frank Ückert, Thomas Bürkle, Cross-institutional data exchange using the clinical document architecture (CDA) , Volume 74, Issues 2–4, *International Journal of Medical Informatics*, March 2005
- [3] Canada Health Infoway, <http://www.infoway-inforoute.ca> accessed Oct 2013
- [4] Singapore's National Electronic Health Record, <http://www.moh.gov.sg>, accessed Oct 2013
- [5] NHS (National Health Service), <http://www.nhscarerecords.nhs.uk/> carerecords, accessed Oct 2013
- [7] Concept of Operations: Relating to the introduction of a Personally Controlled Electronic Health Record System (2011), Australian Department of Health and Ageing
- [8] Implementation Guide for CDA Release 2.0 Consolidated CDA Templates (US Realm). December 2011
- [9] Logical Observation Identifiers Names and Codes (LOINC®), <http://search.loinc.org> accessed Oct 2013
- [10] Clinical Terminology Service , <http://clincle.com> accessed Oct 2013
- [11] IHE IT Infrastructure (ITI) Technical Framework, Revision 9.0,
- [12] Christopher G Chute, MD, DrPH, Jyotishman Pathak, PhD, Guergana K Savova, PhD, Kent R Bailey, PhD, Marshall I Schor, Lacey A Hart, Calvin E Beebe, and Stanley M Huff, MD." The SHARPN Project on Secondary Use of Electronic Medical Record Data: Progress, Plans, and Possibilities" *AMIA AnnuSymp Proc.* 2011; 2011: 248–256