Exploring Complexity in Health: An Interdisciplinary Systems Approach A. Hoerbst et al. (Eds.) © 2016 European Federation for Medical Informatics (EFMI) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/978-1-61499-678-1-297

# A Strategy for Reusing the Data of Electronic Medical Record Systems for Clinical Research

# Yasushi MATSUMURA<sup>a,b,1</sup>, Atsushi HATTORI<sup>b</sup>, Shiro MANABE<sup>a,b</sup>, Tsutomu TSUDA<sup>b</sup>, Toshihiro TAKEDA<sup>a</sup>, Katsuki OKADA<sup>a</sup>, Taizo MURATA<sup>c</sup> and Naoki MIHARA<sup>a</sup>

<sup>a</sup>Department of Medical Informatics, Osaka University Graduate School of Medicine,

Osaka, Japan

<sup>b</sup>MKS Inc., Osaka, Japan

<sup>c</sup>Division of Medical Informatics, Osaka University Hospital

Abstract. There is a great need to reuse data stored in electronic medical records (EMR) databases for clinical research. We previously reported the development of a system in which progress notes and case report forms (CRFs) were simultaneously recorded using a template in the EMR in order to exclude redundant data entry. To make the data collection process more efficient, we are developing a system in which the data originally stored in the EMR database can be populated within a frame in a template. We developed interface plugin modules that retrieve data from the databases of other EMR applications. A universal keyword written in a template master is converted to a local code using a data conversion table, then the objective data is retrieved from the corresponding database. The template element data, which are entered by a template, are stored in the template element database. To retrieve the data entered by other templates, the objective data is designated by the template element code with the template code, or by the concept code if it is written for the element. When the application systems in the EMR generate documents, they also generate a PDF file and a corresponding document profile XML, which includes important data, and send them to the document archive server and the data sharing saver, respectively. In the data sharing server, the data are represented by an item with an item code with a document class code and its value. By linking a concept code to an item identifier, an objective data can be retrieved by designating a concept code. We employed a flexible strategy in which a unique identifier for a hospital is initially attached to all of the data that the hospital generates. The identifier is secondarily linked with concept codes. The data that are not linked with a concept code can also be retrieved using the unique identifier of the hospital. This strategy makes it possible to reuse any of a hospital's data.

Keywords. Electronic medical record, clinical research, template

# 1. Introduction

The data entered into electronic medical records (EMR) are expected to be reused for other purposes, especially for clinical research. If the data needed for a case report form (CRF) can be retrieved from an EMR database, data collection activities become much more efficient [1]. In the present circumstances, an electronic data capture (EDC)

<sup>&</sup>lt;sup>1</sup> Corresponding Author: matumura@hp-info.med.osaka-u.ac.jp

system is used to collect data for clinical research. However, because the EDC system is completely separated from the EMR, a user must transcribe the data from the EMR to the EDC system. We previously reported that we developed a system to simultaneously record an EMR progress note and a case report form (CRF) using a template in the EMR. The electronic CRF was compatible with the Operational Data Model (ODM) produced by CDISC [2], is sent to clinical data management system (CDMS) in a data center [3]. This system can prevent the entry of redundant data into the EMR and CRF, which not only makes the data collection process more efficient but also reduces transcription errors.

If the data originally stored in the EMR database can be retrieved by templates, the data collection process becomes more efficient. For example, in the case that a CRF requires the ejection fraction (EF) measured by an ultrasound cardiogram (UCG), the ability to retrieve the data from the patient's UCG report and automatically populate the EF field in the CRF, would reduce the time and effort required to make the CRF. In this paper, we describe a method for retrieving data from the EMR database using the template.

## 2. Outline of template system

The template is a system that presents frames to users and allows them to enter data according to the frame. We developed a template system that can dynamically change the subsequent frames according to the data that is entered [4]. This was implemented in the EMR system and has been used for years. The template system outputs both structured data (represented in XML [Data XML]) and narrative text data, which are generated by a natural language generation technique. The master data that control the contents of a template are also represented in the XML file (Template XML). The core program of the template (Template engine) reads the Template XML, and then functions as the template. This Template engine was developed using the .NET Framework (version 3.5).

The Template XML file is composed of various units consisting of an item and options or a textbox arrayed in a tree structure. An element code, which is unique in a template, is assigned to each item and option. The concept code is linked to an element code if required. After a user enters data into template, the Template engine creates a Data XML file, which consists of data with an element code, which is linked to the concept code if it exists.

When the Template engine is invoked by each application program, it reads the corresponding Template XML and opens the template. After data are entered using the template, the Template engine outputs the narrative text data, which is returned to the application, and the Data XML file, which is stored in the Data XML database. The Data XML file is transferred and the data elements that are included in the file are stored in the Template element database where one record corresponds to one element.

## 3. Reuse of data in the EMR using the template

The data collected by the template are not only entered by users but also retrieved from the EMR database. Although the databases in the EMR store various kinds of data, the data required by clinical research is limited. The source data for the retrieval of data in the current system are the patient's name, sex, birthday, height and weight, laboratory test data, hospitalization and discharge data, operation data, prescription data, data entered by other templates, and data stored in a data sharing database.



Figure 1. Data flow for Template system. (The italic labels of the arrows explain the process or the content that is being transferred)

We developed interface plugin modules that retrieve data from databases in the EMR. The template engine calls the plugin module to retrieve data according to the designation written in the Template XML file.

#### 3.1. Retrieving data from the databases in the EMR

In the case of retrieving laboratory test data, the patient ID, item code, start date, end date and number of data retrieved are set as the parameters of the interface module. The item code in the database in the EMR is a local code. On the other hand, the item designated in the Template XML file is written using a universal keyword. Thus the Template engine converts the universal keyword to the corresponding local code by referring to a data conversion table. It also receives the multiplied number to convert the values to a single standard unit value. The Template engine will then call the interface module and set the local code as the parameter.

The number of data parameters controls the selection of the candidate data during a designated period. For example, if the number of data is set as 1, it retrieves the data obtained before the end date and the most recent date. If this parameter is not set, all of the data obtained during the designated period are returned. If the number of returned data is more than one, the returned data are passed to the data picking module that allows users pick one of the candidates, which is returned to the Template engine.

The Template engine populates the data with the corresponding item data after multiplying the number obtained from the data conversion system.

#### *3.2. Retrieving data from the template element database*

There are two methods for retrieving template data. One is to designate data using the Template element code with the Template code. Using this method, any elements can be retrieved. The second method is to designate a concept code. The second method is used to retrieve data that may have been entered by several templates.

## 3.3. Retrieving data from the data sharing database

In the EMR system, when an application generates a document that should be stored as medical record, it also generates a PDF file and the corresponding document profile data represented in the XML file and sends them to the integrated document archive server and to the data sharing server [5]. The Document profile XML file includes the patient ID, document class code, event date, and other important data. The important data are represented as items with item codes and values.

Because the document class code is a local code that is unique to the hospital and the item code is unique to the document, the item code and document class code identifies the item element in the hospital. By linking a concept code to this item element identifier, the item element can be retrieved by designating a universally available concept code.

The interface module for the data sharing database searches for the corresponding data using the patient ID and the concept code.



Figure 2. The data flow during the retrieval of data by the template (Bold arrows indicate files, dashed arrows indicate data. The labels in italic text explain the process or the content that is being transferred)

# 4. Discussion

There is a great desire to reuse data entered into EMR databases for clinical research [6]. We developed an interface plugin module for our template system to respond to this request. In order to retrieve objective data, the system needs to acquire a local item code from a universal concept keyword.

Examination report data are among the most frequently required types of data in clinical research. However, the data are usually buried in free text. Retrieving data from text is difficult. Whenever possible, we use templates to write records, especially when the recording data has the potential to be reused in other applications. For example, the reports of the UCG are written using the template. Thus, the ejection fraction can easily be retrieved from the UCG reports in the CRF of a study on cardiac disease. The pathological examination reports in our hospital are generated by a specific system.

This system also generates a PDF file and a document profile in an XML file that includes the diagnosis, histological type, and other details as important item data. These data are stored in a shared database, and can be retrieved in a CRF via the template.

Open EHR has been proposed for generating structured data that can be reused [7]. Open EHR employs a strategy of generating templates using a combination of archetypes, which are originally standardized. The data entered using the open EHR technique can be reused with absolute certainty. On the other hand, time and manpower are necessary to create templates and the data entered using the template may not necessarily be reused. In Osaka University Hospital, we prepared more than 2,000 templates for use in clinical practice. Other hospitals use similar but different templates that are adapted to their clinical practices. We therefore believe that the productivity of the templates should therefore be considered with regard to practical use. We employed a strategy in which unique identifiers were attached to all of the data generated by a hospital, which was secondarily linked with the concept code. The data that is not linked to a concept code can be retrieved using its identifier within the hospital. The data conversion system, which returns a local code corresponding to a universal concept keyword, is helpful for this task. This strategy is flexible and makes it possible to reuse any of the hospital's data.

## Acknowledgement

This research is partially supported by the Project Promoting Clinical Trials for Development of New Drugs and Medical Devices from Japan Agency for Medical Research and Development, AMED.

#### References

- [1] CDISC Electronic Source Data Interchange (eSDI) Group. Leveraging the CDISC standards to facilitate the use of electronic source data within clinical trials version, November 20, 2008. < http://www.cdisc.org/stuff/contentmgr/files/0/2f6eca8f0df7caac5bbd4fadfd76d575/miscdocs/esdi.pdf >
- [2] Clinical Data Interchange Standards Consortium (CDISC) (Internet]). 15907 Two Rivers Cove, Austin, Texas, USA; @2008. Standard available from: http://www.cdisc.org/models/odm/v1.3/index.html (cited Apr 20, 2009).
- [3] Matsumura Y, Hattori A, Manabe S, Takeda T, Takahashi D, Yamamoto Y, Murata T, Mihara N. Interconnection of Electronic Medical Record with Clinical Data Management System by CDISC ODM. *Stud Health Technol Inform*, 205 (2014), 868-72
- [4] Matsumura Y, Kuwata S, Yamamoto Y, Izumi K, Okada Y, Hazumi M, Yoshimoto S, Mineno T, Nagahama M, Fujii A, Takeda H. Template-based data entry for general description in medical records and data transfer to data warehouse for analysis. *Stud Health Technol Inform* 129 (Pt 1) (2007),412-416
- [5] Matsumura Y, Kurabayashi N, Iwasaki T, Sugaya S, Ueda K, Mineno T, Takeda H. A Scheme for Assuring Lifelong Readability in Computer Based Medical Records. *Stud Health Technol Inform* 160 (Pt 1) (2010), 91-95
- [6] El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic PY, Jaulent MC, Daniel C. Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform. Dec;44 Suppl 1(2011), S94-102
- [7] Atalag K1, Yang HY2, Tempero E3, Warren JR4. Evaluation of software maintain ability with open EHR - a comparison of architectures. *Int J Med Inform.* 83(11) (2014), 849-59.