

The Electronic Health Record as a Clinical Study Information Hub

Naoto Kume

EHR Research Unit, Department of
Social Informatics, Graduate School of
Informatics, Kyoto University
/ Kyoto, Japan

kume@kuhp.kyoto-u.ac.jp

Kazuya Okamoto

Division of Medical Information
Technology and Administration Plan-
ning, Kyoto University Hospital
/ Kyoto, Japan

kazuya@kuhp.kyoto-u.ac.jp

Tomohiro Kuroda

Division of Medical Information
Technology and Administration Plan-
ning, Kyoto University Hospital
/ Kyoto, Japan

tomo@kuhp.kyoto-u.ac.jp

Hiroyuki Yoshihara

EHR Research Unit, Department of
Social Informatics, Graduate School of
Informatics, Kyoto University
/ Kyoto, Japan

lob@kuhp.kyoto-u.ac.jp

Abstract

The use of Electronic Health Records (EHRs) is spreading rapidly in several countries. The systems currently used, however, are not designed to permit secondary use of collected data. We present a new design for an EHR system that is capable of connecting information from multiple sites for use in clinical studies by means of an accounting information system and a hospital information system (HIS). This EHR system was designed for healthcare facilities in the Kyoto region. This paper describes how the conventional system can be extended into an EHR system that serves as a clinical information hub.

1 Introduction

EHR is defined by ISO as repository of information regarding the health status of a subject of care, in computer processable form, stored and transmitted securely and accessible by multiple authorized users, having a standardized or commonly agreed logical information model is the support of continuing, efficient and quality integrated health care (ISO 2005). In the United States and New Zealand, an EHR system used by multiple hospitals and clinics in a given region is known as an electronic medical record (EMR) system. Accordingly, this paper distinguishes an EMR system from an EHR system, which is restricted to only one facility.

The original motivation for sharing medical records between facilities was to promote collaboration between clinical facilities. Because medical costs are very high and still rising rapidly worldwide, reducing unnecessary medical tests and duplicate prescriptions is a social necessity. EHR is expected to help accomplish this by improving information sharing. Several countries have established national EHR programs with government support; these include New Zealand, which uses a national healthcare IT plan known as eHealth (National Health Information Technology Board), and Australia, which has established a group (National e-Health Transition Authority) to promote its patient-controlled system. Canada has likewise established a nationwide organization (Canada Health Infoway) to distribute EHR data across all provinces and territories. Finland and Singapore are also promoting the establishment of a national database of medical records.

In most cases, the sharing of medical records is for short-term reference purposes only, and only a document index is exchanged. In Singapore, for instance, secondary use of medical information managed by the public healthcare service (MoHH) is strictly restricted by law. Therefore, the design of Singapore's EHRs is based on document index convergence rather than unified data storage. Since sharing an EHR involves sending and receiving different types of documents, the accumulation of documents can provide a data source extrapolating actual clinical

activities. Note that clinical facilities are not guaranteed organization continuity so that the data repository continuity of each facility is also not guaranteed. Therefore, this study focuses on EHR systems with centered repositories storing different types of medical documents. This study aims to provide a design for an EHR that can serve as an information hub and especially as a resource in gathering data for clinical studies.

2 Methods

Within the last two decades, EMR has been installed in most university hospitals in Japan. These installations were performed by different vendors in different hospitals and regions. Therefore, before medical information can be exchanged, users must define a standard format for data exchange. Because there are so many medical document formats in use, no standard format has yet been agreed upon. Instead, every hospital has its own preferred format. EMR systems can be classified based on structure into two types: a centralized database type and an information locator type. The centralized database type receives test results from a lab test branch system and stores as well as the other medical records such as prescriptions, reports and summaries. The information locator type stores only medical records and pointers to the test results and reports; the body of test results and reports are handled by branch systems. From the viewpoint of EHR construction, an EMR of the information locator type is not cost-efficient because it has to collect documents from every single branch system. On the other hand, an EMR of the centralized database type requires only medical records to be converted to a standard exchange format. In order to support analysis of medical records, an EHR aiming to serve as a source for clinical study data should employ the centralized database model; otherwise, the initial cost and update costs could be prohibitive. In fact, although many Japanese hospitals installed EHR systems during the last two decades, when they were subsidized by the government, most of them were abandoned after the subsidy ended.

If medical records are to be used for research in clinical studies, it is necessary to have access to large numbers of cases, especially for studies on rare diseases. The medical records maintained by a single facility are obviously not sufficient for such studies. When the records in an EHR are used, however, it is hard to avoid controversy over document ownership. Sharing policies vary

from one facility to the next, so that researchers must pay close attention to the policy of each facility, at least until a standard practice for sharing policy develops. Data sharing and the use of large data sets come with certain disadvantages even though obvious benefits.

Clinical studies are performed in strictly controlled environments so that the relation between treatment and outcome can be firmly established. A comparison between treatment as action and clinical outcome as result is mandatory. General action in a hospital can be defined as orders. Orders are recorded in the form of an order history and a description in the medical records in a HIS. In addition, each action is recorded in an accounting system so that the hospital can claim payment for medical services. Order histories in HISs are not typically standardized, but accounting system records are. In fact, because the accounting report format is determined by the government, all facilities in a country use the same one. Accounting records of hospital orders can therefore be used as an action history, allowing comparison of orders across clinical facilities.

Clinical outcomes are recorded in medical records, lab test results, pathology reports, radiology reports and so on. Those records usually consist of structured and non-structured data. Test results are an example of structured data. Pathology and radiology reports, which are described in natural language, are non-structured. Even in hospitals where all outcomes are managed in a structured data format, the structured data format is different from those used by other hospitals. Therefore, a standard structured data format for each item has to be established and used in multiple facilities before outcomes can be compared. In addition, natural language processing (NLP) is necessary to transform non-structured data into ontology components. Some studies have applied NLP manually to certain document types such as the discharge summary. As this is labor-intensive, however, it is not cost-efficient and therefore not an option for most facilities. Therefore, semi-automatic analysis is required.

Overall, the authors defined the requirements for an EHR intended as a data source for clinical studies as follows:

- Centralized database collecting data capable of traversal query among multi-facility records
- Access control to maintain data-source-facility's sharing policy

- Convergence of accounting information and medical records as action and result
- Automatic data mapping engine using massive medical records

Figure 1 illustrates a proposed design for an EHR generating structured data using its own dataset as well as structure mapping between several facilities' data.

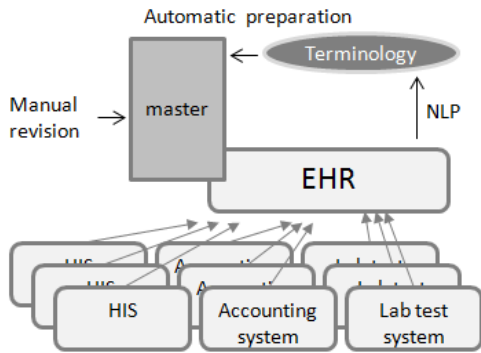


Figure 1. An EHR design promoting convergence of HIS information and accounting information.

If traversal master data is manually maintained, data analysis cannot be completed because of the rate at which new data arrives from other facilities. Therefore, both master data maintenance and medical record exchange requires NLP to achieve semiautomatic data preparation.

So far, there are several obstacles to the realization of this proposed EHR, including data ownership control issues, legal restrictions on data location, the need for NLP technique for automatically generating terminology, the need for a master data definition, and so on. Because we cannot fully implement a real EHR of this kind, therefore, the authors implemented a database to verify how the EHR equipping a centralized database would contribute to a clinical study. Specifically, the authors implemented a clinical study database composed of datasets from several university hospitals. The centralized database constitutes a traversal search environment for accounting information and test results.

There are six requirements of an EHR database suitable for clinical use: available data range, access control, sharing control, search query performance, usability, and database management policy agreement. Available data range depends on the NLP technique as well as the reachable dataset in the HIS. Data export from the HIS to the EHR depends on the conventional hospital setup and policy. Exported data is

aligned in a semi-structured format such as XML. Access control must be extended to researchers as well as medical caregivers and patients. Also, for privacy reasons, results should consist of overviews and abstractive information instead of patient-specific information. Sharing control should be given to each participating hospital's administration; otherwise, the hospital's internal council will hardly be convinced. The database should also be fast enough to allow for a traversal search of multiple datasets that are each constantly growing. To enable such fast searching, each dataset should be optimized for a search query created by a manager who is capable of setting up each researcher's required database query. Because it is too hard for researchers to correctly understand this data retrieved from multiple datasets, a search query manager should be assigned at the datacenter. Finally, an audit council should be organized to reach agreement on any database management issues.

Because of these restrictions, it would be difficult to implement the EHR as designed from scratch. Therefore, as a proof-of-concept, the authors implemented a system that meets the above requirements but is based on a currently available technique and dataset. There are two convergent datasets, a set of accounting information and a set of test results, as action and result information. Four university hospitals contributed to the convergent datasets: Kyoto University Hospital, Chiba University Hospital, Osaka University Hospital, and Miyazaki University Hospital. Although the use of these datasets was approved, the physical setup of a unified database was not allowed. Therefore, the authors installed a virtually centralized database based on a database (Cache). Figure 2 illustrates a database. Each hospital has own dataset in the facility. A virtual datacenter is allowed to access all database by a search query.

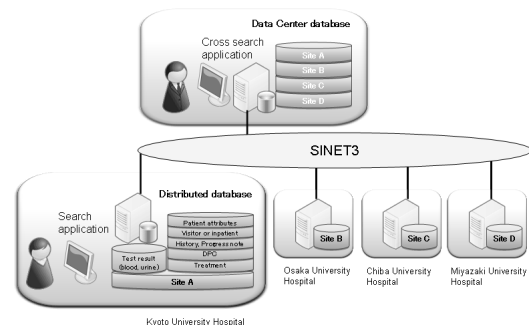


Figure 2. Virtual datacenter connecting four university hospital databases.

A traversal search query is implemented at the datacenter, then equally distributed to all facilities so that each facility can see the search results. Here, if a facility declines to share a search result, an administrator at a local site can stop sharing that result. Also, each search query is discussed and agreed upon by a council consisting of four university administrators beforehand.

So far, there is no proper standard format for test results and no semi-automatic technique for master preparation, so the authors manually prepared a unified master based on JLAC10 and composed of four university test result masters. The accounting information master used here is well established because the format for medical treatment fee claims is almost universal in all hospitals in Japan.

3 Results

The authors carried out five studies using the database. A query manager created a traversal search query by making an optimal search query of each site database. Because the unified test result master was manually maintained, only 300 test result items were available for query, even though over 3000 items are available in each site. The script list as a researcher would see it is presented as Figure 3. The researcher's request was analyzed by the manager and translated to a database query beforehand. The authors proceeded with five traversal queries as follows: Zyvox-treated patients, nicotine addiction treatment patients, teicoplanin-treated patients, bortezomib-treated patients, and influenza patients. Here, the definition of patient is different between facilities because of differences in employed drugs, drug names and applied disease names. Therefore, the query manager was required to find proper combinations of those parameters optimized to each facility's database.

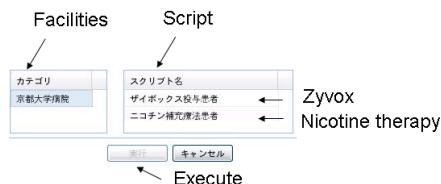


Figure 3. User interface of traversal search query script.

Figure 4 illustrates the result of a search of patient findings. The result can be evaluated by a site administrator before it is shared. If the result is allowed to be shared, the list is accumulated to a statistical result.

No.	施設名	患者ID	性別	生年	処方開始	処方終了	処方開始	処方終了	処方開始	処方終了
1	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-29	2009-02-05	41	
2	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-27	2009-01-28	41	
3	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-28	2009-01-29	41	
4	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-24	2009-01-24	41	
5	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-29	2009-01-29	41	
6	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-26	2009-01-27	41	
7	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-26	2009-01-23	41	
8	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-22	2009-01-23	41	
9	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
10	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
11	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
12	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
13	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
14	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
15	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
16	京都大学病院	E448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	
17	京都大学病院	F448E05800	男	1968	2009-01-22	2009-02-05	2009-01-23	2009-01-23	41	

Figure 4. Traversal search result list for a site administrator.

The virtually centralized database, handling 180 million records including data from 2009 and 2010, completed each query in less than 10 seconds. It was concluded that this level of information processing performance is enough to satisfy usability.

The results show that the number of patients can be compared between hospitals. In other words, a centralized database like this can be beneficial for case finding, especially for finding rare diseases and common treatment procedures.

General patient information such as height, weight, medical questionnaire responses, disease history, vital information, and contraindicated medicines are necessary and must be added to the database, according to a pharmacoepidemiologist.

4 Discussion

Clinical studies require strict control of data to verify the results. On the other hand, epidemiology requires a massive dataset to analyze general information statistically, even if the accuracy of the dataset items is not verified. Therefore, the proposed EHR would contribute to epidemiology as well as case finding in clinical studies. The data sharing policy should begin with the patient's ownership of his or her own data, for no unified database can be achieved based on caregiver's ownership. Also, a semi-automatic data alignment technique to maintain master data and analyze unstructured documents is necessary. NLP would be convoluted to the data cleansing cycle.

5 Conclusion

This paper proposed an EHR designed to serve as an information hub for clinical studies. A centralized EHR database was defined to achieve

traversal medical record search of multiple facilities. The authors implemented a virtually centralized database as a proof-of-concept. The database contained accounting information and test results. Five case studies were performed to find patients from multiple facilities. The authors concluded that the database cannot be used directly in clinical studies but is beneficial in case finding as well as epidemiologic analysis.

References

- Canada Health Infoway, <https://www.infoway-inforoute.ca/> (Last access: Jul 20, 2013).
- Cache, InterSystems Corporation,
<http://www.intersystems.com/cache/index.html>
(Last access: Jul 20, 2013).
- ISO 2005, "Health informatics — Electronic health record — Definition, scope and context", TECHNICAL REPORT ISO/TR 20514:2005(E), p. 2, 2005.
- MoH Holdings in Singapore,
http://www.mohh.com.sg/about_mohh.html (Last access: Jul 20, 2013).
- National Health Information Technology Board,
<http://www.ithealthboard.health.nz/> (Last access: Jul 20, 2013).
- National E-Health Transition Authority,
<http://www.nehta.gov.au/> (Last access: Jul 20, 2013).