

A Framework for Integration and Standardization of Data from Heterogeneous Sources

C. Sunil Kumar^{1,*}, C. V. Guru Rao², A. Govardhan³

¹CSE Department, ACE Engineering College, Hyderabad, (Affiliated to JNTU University), 501301, India

²CSE Department, S.R Engineering College, Warangal, (Affiliated to JNTU University), 506371, India

³School of IT, JNTUH, Hyderabad, (Affiliated to JNTU University), 500085, India

Abstract It is clear that in today's organizations, new and existing applications require access to data stored in several pre-existing databases held at several local and remote locations. Therefore, a main criterion required by most complex organizations, is the provision of collaboration possibilities and information integration mechanisms among distributed, heterogeneous, and autonomous database systems. The development of an application provides interoperability and information integration among distributed systems, via the deployment of database standards and emerging Internet technologies. It is one of the most challenging approaches in the area of integrating heterogeneous information from autonomous sites. In this context, the work described in this paper focuses on the design and development of a Generic Information Exchange (GIE) System. The system supports a wide variety of applications with efficient means for their interconnection and interoperation, while preserving their heterogeneity, distribution, and full autonomy. An example of the interoperability problem is found in the healthcare domain, where each hospital, or even each department in a hospital, maintains its own database. In this environment it is very important to permit users to locate and access data from several remote databases, supporting the needs of patient care, daily operations of the hospitals and research consultations. It necessitates the sharing and exchange of data related to clinical, administrative, managerial and research (statistical) information.

Keywords EHR, GIE, XDB, DBX, HL7, Database

1. Introduction

The design and development processes of advanced applications in scientific and system engineering domains consider different data modelling and Information management strategies. Data models define the data structures and relationships among the data, to reflect the proper representation of the information each application needs. The information management strategies however, depend on the global architecture design and the chosen database system to fulfil the functionalities required by the application.

Diversity of the used information management approaches is usually due to different characteristics and requirements of each application. Due to the complex requirements of emerging applications, several scientific and business oriented organizations from biology, medicine, physics, astronomy, engineering, e-commerce, etc. have realized the need to reconsider their information management systems towards better addressing of collaborative work. Therefore, these organizations are required to provide appropriate products and services, and to better react to the new

information management requirements in terms of data integration.

1.1. Problem Statement

Traditionally, information integration and data translation among different heterogeneous and autonomous sites were considered a completely manual process, where either the user or the database administrator must do the data translation and exchange. Nowadays, the problem of data integration and information exchange among heterogeneous data sources has become a challenging issue to be studied, and different integration approaches are being examined and evaluated.

This paper addresses the issue of information integration and standardization for systems interoperation among heterogeneous and autonomous applications, and mainly addresses solutions related to the requirements for interoperation among different healthcare systems in term of information exchange and services[11].

Healthcare systems have grown rapidly in the last decade. They moved from isolated software systems in primary health centres towards solutions which support a continuous medical process. It includes multiple healthcare professionals, institutions; utilize ubiquitous computing healthcare environments with technological advances. In an interactive environment, there is a need to look at the information

* Corresponding author:

ccharupalli@gmail.com (C.Sunil Kumar)

Published online at <http://journal.sapub.org/database>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

sharing amongst healthcare systems. Healthcare professionals whenever they require all the relevant patient medical data in an appropriate format must be available to them. The prediction is that in future there would be an interoperable information exchange in the form of electronic health record (EHR). The research will address the current challenges and the benefits of the interoperability among healthcare systems.

The Proposed method explores a framework for interoperable healthcare systems in general and it attempts to create a successful and workable solution. The general solution of the framework is:

- i. Agree on a standard data object structure that will serve as the intermediate among hospitals.
- ii. Exchange XML based document
- iii. Map this document to the standard data object structure
- iv. Receive and store the document
- v. Map data to existing relational database attributes

Organization of the Work

Following, is a brief description of the contents of each chapter in the thesis:

Section 1: Introduction

This Section gives a general overview of the research, problem statement, presents the research issues and describes the main structure of the paper.

Section 2: Literature Survey

This Section defines the fundamental concepts related to data integration issues. It reviews the interoperable standards, and identifies the required elements to consider for the integration of healthcare data.

Section 3: Proposed Method

In this Section a GIE System is proposed for supporting the integration and standardization of heterogeneous sources.

Section 4: Experimentation

This Section describes a case study on healthcare systems using proposed XML-to-Database (XDB) mapping and Database-to-XML (DBX) mapping algorithms and test cases in distinct healthcare systems that can exchange healthcare data for interoperability.

Section 5: Results

The results on this system when tested on distinct healthcare systems are illustrated and analysed.

Section 6: Conclusion

The Section 6 concludes this research work and states the scope for further work. Thesis is ended with bibliography.

2. Literature Survey

In any domain the need for integration is justified in that it facilitates sharing of information between systems and across organizations. The integration of information systems enables smoother coordination and control of organizational processes and delivery. The pressure for tighter integration in any domain results from existence of abundance of different information systems (IS) which is the mirror image of the enormous variation. The several dimensions are: level

(hierarchically organized), geography (municipalities, counties, districts, nations and regions), professional groups, agencies and specialization presents some of the relevant aspects that pushed the evolution of information systems and relevant technologies in order to appropriately support organizations which includes technological advances in systems structure and communications, facilitating the implementation of integrated networks.

In the area of integrating heterogeneous and distributed information sources, the information integration generally implies uniform and transparent access to data managed by multiple databases. The task of an integrated database system is to answer queries that may require extracting and combining data from multiple local/remote data sources.

In an interactive environment, there is a need to look at the information sharing amongst various information systems (For E.g. Banking, Military Services and Health care). The specific situation is characterized by:

- i. Heterogeneities in hardware and software solutions.
- ii. Heterogeneities in the structure, purpose and deployment.

The major problems of healthcare system globally in the 21st century have to do with the quality, safety, effectiveness, accessibility, Standards(e.g. HL7) and cost[1, 2, 3, 4 and 5]. Regardless of economic and social status, all nationals are looking for safe, quality and effective healthcare services at a reasonable cost. The uses of ICT in healthcare primarily aim to lessen the healthcare problems for better health care. Although there have been failed implementations in the past, the evidence shows that health ICT systems help to establish a quality in the healthcare industries[6].

Implementation of interoperable healthcare system may give rise to various challenges such as introducing EHR standards[7]. The barrier to achieve interoperable healthcare system is the enormous cost needed to re-establish a complete medical system. It includes a need to update, changes in software or hardware, training. All of them are sophisticated. However, standardization is the major step required for sharing and classifying healthcare data.

3. Proposed Method

A GIE System is proposed with a standards-based messaging engine, which is event-driven and provides foundational services for more complex software systems. It is both operating system and programming language independent and provides interoperability between different platforms[8]; for example, between VB, Java and .NET applications. It is not necessarily Web-Services based (although this is often the case) and uses XML as a standard communications language[9]. It provides communication between disparate information systems which can all connect to it regardless of the type of software or hardware used by adopting the following techniques as follows:

3.1. Techniques Adopted

3.1.1. Health Level 7 Version 3.0

The term HL7 stands for "Health Level Seven" and it refers to the top seventh level of the ISO communications model. The application level addresses the definition of the data to be exchanged, the timing of the exchange, and the communication of definite errors to the application. Specifically, to create flexible, cost effective approaches, standards, rules, methodologies for interoperability between healthcare systems. The core classes of the HL7 V3 is shown in the Figure 3.1

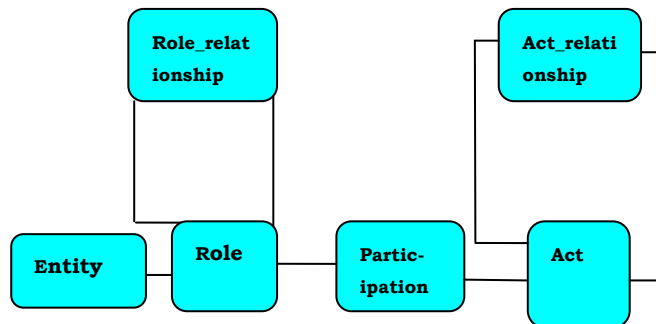


Figure 3.1. The core classes of the HL7 RIM

The "Act class" represents all the actions and happenings analogous to a verb to be documented through the healthcare process.

The terms 'Act', 'Action', and 'Activity' are all used interchangeably. It captures all the events that have happened in the past, that are currently happening or that are expected to happen in the future.

The "Entity class" represents any physical thing or being analogous to nouns that takes part or is of interest in the healthcare. Although it instantiates any physical thing or group of physical things including living subjects and organisms, it does not include the roles that things can play or the acts that things can perform.

The "Role class" ties an entity to the acts that it plays or provides, specifying how a particular entity participates in a particular act. At the same time, it connects the Entity playing the "Role", to the specified "Act", thus expressing the context for the Act in terms of who performed it.

The "RoleLink" class specifies the connections and dependencies that exist between two different and individual role objects. The Participation class specifies a relationship between a particular Role instance and a particular Act instance. Such relationships include "Act to Act" associations, as well as "Source/Target" associations between the objects.

The "ActRelationship" class associates a pair of Act objects, representing a connection from one Act to another one.

The Acts connect to Entities in their Roles through Participations. It also connects to other Acts through ActRelationship. Examples for each core class of the RIM model are presented in the Table 3.1.

3.1.2. XML – The Fast Track to Information Exchange

XML's greatest advantage is that it is a user-driven, open standard for exchanging data both over corporate networks

and between different enterprises, notably over the Internet. XML transports the Meta data (the information about the data) together with the relevant data, thus allowing its meaning to be easily interpreted. In addition, XML enables suitably coded documents to be read and understood without difficulty by both humans and machines. XML as a data interchange format is compelling, primarily because it gives developers:

i. A language with which to more easily identify interoperability problems.

ii. A common syntax and tool set with which to fix them.

The proposed method explores a framework for integration and standardization of data from heterogeneous sources, in general and it attempts to create a successful and workable solution with the proposed algorithms.

Table 3.1. RIM backbone classes examples

Class Name	Definition	Example
Act	action that is being done, has been done, can be done	clinical observation, discharging a patient
Entity	physical thing or group of physical things participating in Acts	Person, medical device
Role	competency of an entity participating in an Act	Patient, nurse, doctor
Participation	Association between an Act and Role, with entity playing that Role	Dr.Shiva prescribes a therapy for patient ram
ActRelationship	Direct association between source Act and target Act	A biopsy procedure as a result of an observation
RoleLink	Connection between two Roles	Medical Expert - Hospital

3.2. Proposed Algorithms

3.2.1. DBX Algorithm

The DBX algorithm extracts data from the database. It generates XML document and allows an exchange in any domain.

Step1: Connects to the database and performs a select query to get the data (Attributes).

Step2: Creating a new XML DOM document tree 'T', in which the data will transfer to it.

Step3: The first element created in the XML document is called the "root" element.

Step 4: For each row: add a new element to the XML document, using the table name, then insert it into the document as a child of the root element.

Step 5: Loop through each column in the current row, and insert the field name, and corresponding value.

Step 6: Create a new element for the field and then insert it as a child to the current database row.

Step 7: Add the field value as a text node, and then insert it as a child element to the current field node.

Step 8: Repeat from Step 4: These loops do not terminate until they have processed every column of every row which has been retrieved from the database.

Step 9: Returns the completed XML document as a string.
 Step 10: Insert the results into the XML document
 Step 11: End.

3.2.2. XDB Algorithm

The XDB algorithm takes XML document as input and maps its contents to target database attributes.
 Step1: Read the EHR Document (XML File) as String.
 Step 2: The messages are parsed in the tree 'T'.
 Step 3: The first element found in the XML document is called the "Root" node.
 Step 4: Find child nodes in the XML document of the root node.
 Step 5: For each row: read and count the no. of elements of XML document, in tree 'T' to construct RIM object.
 Step 6: Loop through each node finds the corresponding attribute name and its value.
 Step 7: Read map settings from database and place it in a queue.
 Step 8: Place data in a temporary application table.
 Step 9: Take 'n' no of variables which are equal to data base fields count.
 Step 10: Check for the target field and assign appropriate variable with the value.
 Step 11: Creating connection to the database to map XML data.
 Step 12: Repeat from Step 5: These loops do not terminate until all nodes are processed in the document.
 Step 13: End.

4. Experimentation

The GIE System is experimented on distinct healthcare information systems for standardizing the healthcare data. The detailed case study is discussed in the following sub-sections.

4.1. Case study on Healthcare Systems

Healthcare systems have grown rapidly in the last decade. They moved from isolated software systems in primary health centres towards solutions which support a continuous medical process. It includes multiple healthcare professionals, institutions; utilize ubiquitous computing healthcare environments with technological advances. In an interactive environment, there is a need to look at the information sharing amongst healthcare systems.

Healthcare professionals whenever they require all the relevant patient medical data in an appropriate format must be available to them. This improves the quality of the healthcare. In any hospital, the electronic health record (EHR) is unique and it is generated for every citizen. It means that the structure of EHR and the methods used for exchanging their content may vary significantly. This becomes an obstacle for sharing health data or health records between healthcare systems.

Currently, there is no data-level standard which is defined for the storage and retrieval of clinical information within EHR. Most standards organizations, including health level seven (HL7), have emphasized the structure of the messages being exchanged between healthcare systems. It allows significant variation in the content and internal organization of data within that structure. The lack of standardization, particularly of quantitative data, hinders interoperable use and requires a great deal of work on translations among the internal representations which can be transmitted to and understood by another healthcare system.

Interoperability concentrates on the necessity to link up healthcare data. Its goal is to provide health data 24/7 from any healthcare institute. This communication will improve accessibility of the patient record, so clinicians who require patient's demographic or medical information are not bound by the limitations of time or site.

The proposed GIE System is implemented on complex healthcare information systems to provide foundational services. It is based on HL7[1] standard for the exchange, management and integration of health data to generate EHR. It adopts XML to serve as a messaging syntax. The electronic health record is a collection of documents pertaining to a patient from different healthcare sites that are generated at different times. Medical-technical documents like x-rays or cardiograms are appended to the patient health record.

4.2. Creation of Electronic Health Record

The key capability of EHR is to create a single patient-centric view of entire health data captured for his life-time. The creation of EHR requires the components as shown in Figure 4.1. Creating these components is based on standards such as HL7 that allow disparate healthcare systems to communicate. In this process the records need to be retrieved from multiple sites that are part of the patient's healthcare continuum using the above mentioned standards. The next sub-section describes the interoperable EHR.

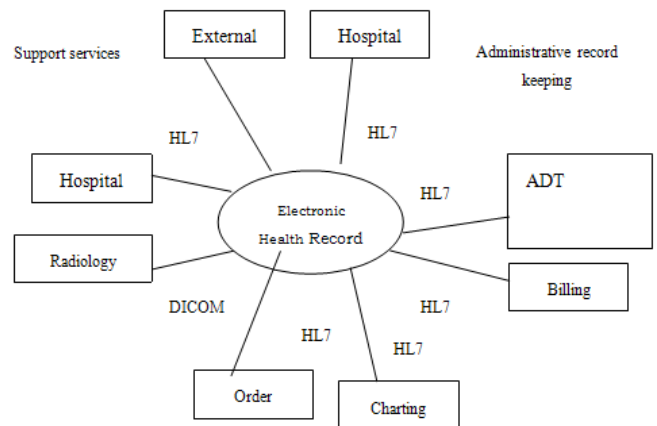


Figure 4.1. Creation of Electronic Health Record

4.3. Access to Electronic Health Record

Every EHR that is created must make a provision to give access to the data stored. The EHR data will be parsed in order to access. The processes required to provide interoper-

erable EHR are proposed as three. They are message generation process, transport process and receiving process as shown in Figure 4.2 and are proposed as a message exchange model. It allows exchanging EHR data as XML document and vice-versa.

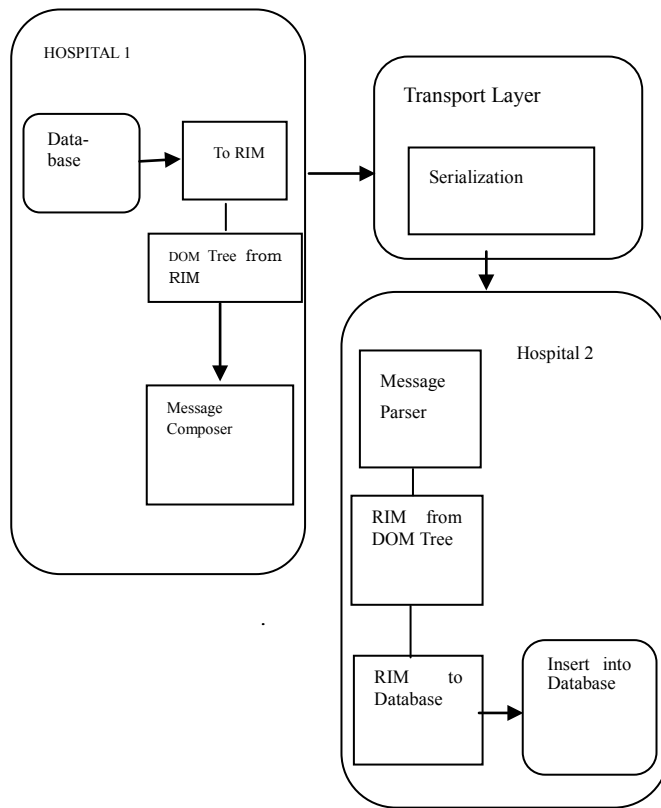


Figure 4.2. Message Exchanging Model for interoperable EHR

The first process is to refer a database, convert to reference information model (RIM), making document object model (DOM) tree and create a message. To start, it references database(s) and retrieves data to form a RIM object. These generated RIM objects are represented over a DOM tree and then composed as XML message.

Example:

The XML based EHR document 'D' is modelled as a DOM object. Document tree 'T', in which nodes represent XML elements and edges represent parent-child relationships between XML elements. For each XML element node, the following notations are used:

- E.Name: the name of the XML element.
- E.Parent: the parent node of E, if root node NULL of T.
- E.Children: the set of child nodes of E denoted by: E.c1, ..., E.cm.
- E.Attributes: the set of XML attributes of E denoted by: E.a1... E.an. The attribute names are denoted by E.ai.Name where $i = 1, \dots, n$.
- E.Value: the values of the set of XML attributes of E are denoted by E.ai.Value, if a non-leaf node value is NULL.

An XML tree for an instance of XML document is illustrated in Figure 4.3. The Patient Health Record node at the top of the tree has two children such as (Patient_data,

Clinical_test). The Patient data has children like Name, Age, and Ref_Hosp and ClinicalTest has children like Pat-Id, Type and Valuation. All these leaf nodes are represented by their respective values and all the values compose a XML message.

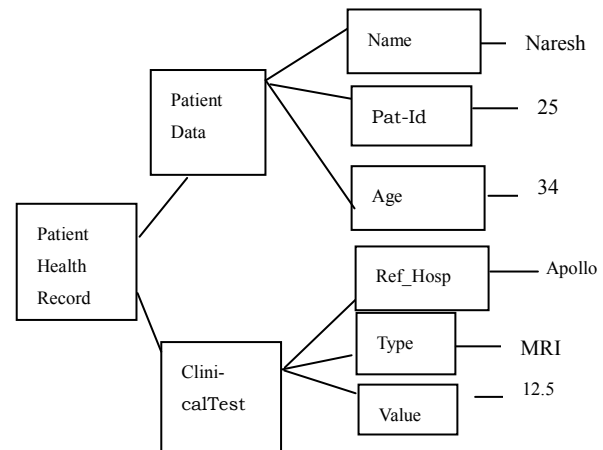


Figure 4.3. Tree representation for the health record

The second process is proposed to transport a message. The aim is to send a message to receiver safely. The XML message is sent by using some transport mechanisms such as e-mail, HTTP, TCP/IP & SOAP.

The final process is to receive the message which is shown in Figure 4.4. The receiving process is to parse, interpret the DOM tree and retrieve the data from a RIM object to store it into database.

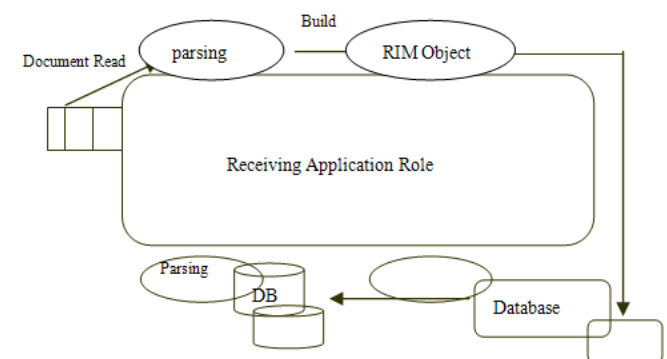


Figure 4.4. Message parsing and mapping fields to the database

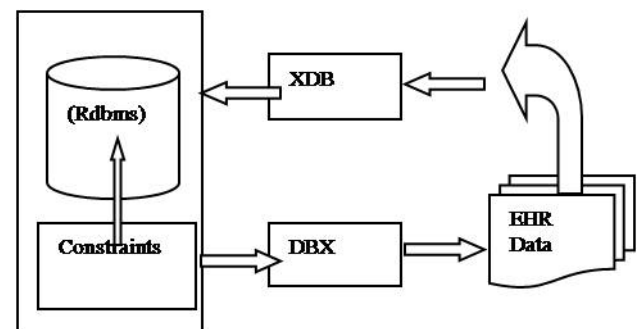


Figure 4.5. XDB and DBX algorithms are used independently or jointly

To create and provide access to interoperable EHR a GIE System is proposed. The Figure 4.5 illuminates the need of

proposal and the algorithms to be employed. To incorporate the EHR data into relational database a XDB mapping algorithm is proposed. The DBX mapping algorithm is proposed to extract health data in XML format from database(s). An interface is in need to integrate these two algorithms as a GIE System in order to support creation and access of interoperable EHR. These proposed algorithms are described in the subsections.

The algorithms which were proposed in previous section are experimented in healthcare systems for interoperability.

4.4. Proposed DBX Algorithm

The DBX algorithm extracts health data from the database. It generates XML based EHR document and allows an exchange. The XML generated EHR data are shown below

```
<? Xml version="1.0" ?>
<? Xml version="1.0" ?>
<Root> <PatientHealthRecord>
  <Clinical_test >
    <Pat-Id> 10000 </Pat-Id>
    <Type> MRI</Type>
  <Valuation> metastasis</Valuation>
  <Clinical_test>
</PatientHealthRecord> </Root>
```

4.5. Proposed XDB Algorithm

The XDB algorithm takes XML based EHR document as input and maps its contents to target relational database attributes. The XDB algorithm tries to incorporate the EHR document elements into database as shown in Table 4.1.

Table 4.1. Element Table constructed by XDB algorithm

Elm-Id	Elm-Name	Elm-Value	Parent-Id
1	XML	Null	Null
2	PatientHealthRecord	Null	1
3	Patient_data	Null	2
4	Name	Naresh	3
5	Age	25	3
6	Ref Hosp	Apollo	3
7	Clinical test	Null	2
8	Pat-Id	10000	7
9	Type	MRI	7
10	Valuation	Metastasis	7

5. Results

The GIE System for healthcare is aimed to provide clinicians with accurate, real-time patient information from geographically disparate locations for making clinical decisions. A combination of paper based and computerized health record systems still exists due to the challenges involved in capturing complex clinical information and costs. In order to make the best possible clinical decision the patients and clinicians are keen to encourage the sharing of health information. It will forbid the duplication of investigations. Sometimes incompatible computer systems prevent sharing of valuable clinical information consumes time to

prepare reports and enormous space to store.

Another issue of focus in paper based record of health information written in free style by most of the clinicians is either not legible or there is a chance to miss/forget some important information. This might lead to serious consequences on patient's health care. Paper based record is a hard copy that can be accessed only by one person at a time and in a given place; it needs physical transfer if required by another person to access at other place. Retrieving requested record from the archive will be a matter of luck, missing a record in such archive is not surprising. In addition, the paper based record gets diminished because of ageing. Sometimes fire accidents or natural catastrophe like floods and earthquakes can completely ruin the archives.

All above mentioned de-merits can be resolved by GIE System. It solves problems of consuming large space and resources for maintaining these paper based records. Most healthcare providers lack the knowledge of medical informatics, standardization and healthcare systems that are not interoperable. The GIE System generates the EHR

related to a patient that contains the patient medical history, routine examinations and findings. This EHR can be shared among healthcare providers.

5.1. Comparison of Results

The results of GIE System when implemented in multiple hospitals are compared with existing hospitals in the following Table 3. The important factors to rate the successful implementation of GIE System are listed in the first column of Table 5.1.

The GIE System implemented in multiple hospitals is free from system failures. It is integrated with standards based integration (HL7) for creation and exchange of EHR data had resulted in no conflicts among providers in the fourth factor. The fifth factor describes the training and education factor at the right times for all actors involved in IE. The sixth factor describes the medical experience of handling EHR in healthcare institutions. The last factor describes the sharing of EHR data among caregivers with rated GIE System. The results are self explanatory in the table to demonstrate success of the GIE System.

The Figure 5.1 shows the bar chart between the number of doctors participated in the implementation process (i.e. Maximum number is 30) are represented in Y-axis. The X-axis represents the views of doctors on benefits of implementation by way of their usefulness. The outcome of implementation leads towards responses, research & planning and improved treatment. It is disclosed when all doctors are involved in implementation of interoperable EHR which tremendously improve quality care.

The GIE System enables the seamless transfer of electronic information and images (DICOM) standard without the need to replace existing systems. By ensuring the timely and reliable delivery of information, GIE System helps facilitate the improvement of patient care while reducing costs.

The performance and time complexity of the GIE System is analyzed in several scenarios and found scalable, accurate.

The time complexity of algorithm XDB is $O(n^3)$ where n is the number of elements and attributes in EHR document. It is clear that, the first for loop statement in line 9 will be executed for n_1 times, where n_1 is the number of elements in T , including the document element. XDB have $n_1 < n$. Therefore, all the operations involved in for loop spend constant amount of time. Hence, it is clear that the XDB algorithm runs in $O(n^3)$ time complexity.

Table 5.1. Comparison of results

Factors important for implementation	Karolinska EHR Implementation (Within Hospital)	Sankara Nethralaya EHR Implementation (Within Hospital)	Interoperable EHR implementation through GIE System (Multiple Hospitals)
User friendly, efficiency in use and accessibility	Yes	Yes	Yes
Health staff acceptance and implementers responses	Yes	Yes	Yes
System failures	No	No (occasionally yes)	No
No conflicting suitability	Yes	Yes	Yes
Training and education provided at the right times, amount and quality	Yes	Yes	Yes
Previous computer or EHR experience	Yes	No (not with this EHR system)	Yes
Interoperable EHR among healthcare providers	No	No	Yes

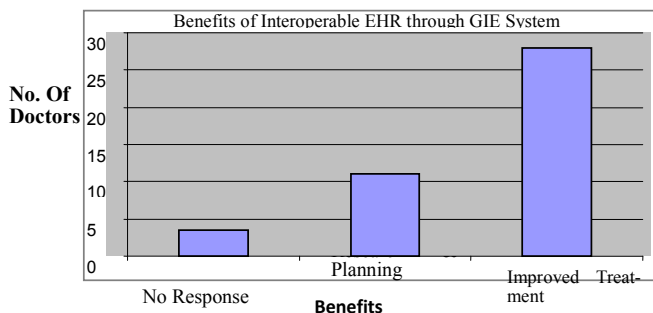


Figure 5.1. Bar chart for Interoperable EHR with GIE System

6. Conclusions

This investigation exposed a number of issues that have come up in all interoperability scenarios of healthcare systems. Interoperability can be successful when there is some level of coordination and communication in the exchange of the healthcare information among the healthcare providers with authentication and authorization. The above solution works for creation of patient's health record which is interoperable among healthcare systems. It is a good starting place for interoperable exchange of healthcare information resulting in quality care of patients. GIE system developed is robust, reliable, secured and scalable.

ACKNOWLEDGEMENTS

The author would like to express their sincere gratitude to the Management of ACE Engineering College, Hyderabad for their constant encouragement and co-operation.

REFERENCES

- [1] Health Level Standard version 7 retrieved May 12, 2007 from "www.hl7.org"
- [2] World Health Organization, "Building foundations for eHealth: progress of member states: report of the Global Observatory for eHealth. Geneva", WHO Press, 2006.
- [3] Chaudhry et. al. "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care", Annals of Internal Medicine, 144(10), 2006, pp.742- 52.
- [4] Stolberg H.O. "The Canadian Health Care System: Past, Present, and Future", Journal of the American College of Radiology, 1, 2004, pp.659-670.
- [5] Tang P.C. "Key Capabilities of an Electronic Health Record System", Institute of Medicine, 2003.
- [6] Dick S.R et al. "The Computer-Based Patient Record: An Essential Technology for Health Care, Revised Edition", National Academy Press, 2004 .
- [7] Bossen. C. "A National Standard for Electronic Health Records, Computer supported cooperative work", Canada, 2006, pp.69-78
- [8] C.Sunil Kumar et al. "Interoperability And Assessment of RIS with other legacy systems", IETECH Journal of Information Systems, 2008, Vol-II, No.1, Pages 006-011.
- [9] C.Sunil Kumar et al. "Usage of XML Technology in Electronic Health Record for Effective Heterogeneous systems integration in HealthCare", International Journal of Medical Engineering and Informatics, Vol.1, No.4, 2009, Pages 399-405