

PACS and Electronic Health Records

Simona Cohen, Flora Gilboa^{*}, Uri Shani

IBM Haifa Research Labs, Haifa University, Mount Carmel, Haifa 31905, Israel

ABSTRACT

Electronic Health Record (EHR) is a major component of the health informatics domain. An important part of the EHR is the medical images obtained over a patient's lifetime and stored in diverse PACS. The vision presented in this paper is that future medical information systems will convert data from various medical sources – including diverse modalities, PACS, HIS, CIS, RIS, and proprietary systems – to HL7 standard XML documents. Then, the various documents are indexed and compiled to EHRs, upon which complex queries can be posed. We describe the conversion of data retrieved from PACS systems through DICOM to HL7 standard XML documents. This enables the EHR system to answer queries such as “Get all chest images of patients at the age of 20-30, that have blood type ‘A’ and are allergic to pine trees”, which a single PACS cannot answer. The integration of data from multiple sources makes our approach capable of delivering such answers. It enables the correlation of medical, demographic, clinical, and even genetic information. In addition, by fully indexing all the tagged data in DICOM objects, it becomes possible to offer access to huge amounts of valuable data, which can be better exploited in the specific radiology domain.

Keywords: Electronic Health Record (EHR), PACS, DICOM query, HL7, Clinical Document Architecture (CDA), XML

1. INTRODUCTION

An Electronic Health Record (EHR) is defined as digitally stored healthcare information throughout an individual's lifetime with the purpose of supporting continuity of care, education, and research. An EHR may include such things as observations, laboratory tests, medical images, treatments, therapies, drugs administered, patient identifying information, legal permissions, and so on. EHR systems will contribute to more effective and efficient patient care by facilitating the unification of clinical information from large populations of patients across care sites. Today, patient clinical information is generally stored where it was created, and is not always available at point-of-care. Future EHR systems will facilitate electronic access to all the patient's data over his lifetime, from any point-of-care, and through various protocols and interfaces. It will enable telemedicine between a patient and professionals, or teleconsultations among professionals, as well as self-health management via the internet, and clinical trials management. Transferring patient information automatically between care sites will speed delivery and reduce duplicate testing and prescribing. Automatic reminders will reduce errors, improve productivity, and benefit patient care.

The human genome has been sequenced and the research community is working on correlating patient diseases with the patient's genes to discover new treatments and develop personalized medications. The vision is that the EHR system will include the patient's genomic information to facilitate genomic-based diagnosis and personalized treatments. The research community is working on the challenge of encoding DNA, proteins, and polymorphism features in the EHR. Medicine of the future will be based on EHRs, and the infrastructure of EHR systems should be dynamic and flexible to adapt to the various types of medical data.

^{*} Correspondence: <mailto:flora@il.ibm.com>; IBM Labs, Haifa University, Mount Carmel, Haifa, 31905; phone: +972 4 8296340, fax: +972 4 8296116

Many organizations are interested in anonymous mining of medical data and correlating the various medical parameters could lead to new conclusions and innovations. The research community would like to investigate and analyze the vast amount of existing clinical data, and find correlations between various diseases and medications, protocols and survival status, diseases and genomic, and so forth. Pharmaceutical companies would like to mine the medical data for drug discovery and development. The Food and Drug Administration (FDA) could mine the data for the process of approving new drugs. Insurance companies could provide improved personalized health insurance policies. Government bodies and world healthcare organizations could enhance their observations by correlating environmental information and medical data. EHR systems with extensive indexing and query capabilities, together with an appropriate security model, will enable such anonymous mining of medical data. In particular, the EHR system will enable mining the various medical images taken on a patient over his lifetime and stored in diverse systems. This paper focuses on combining the patient images in the EHR system.

Medical images have traditionally been analogical and stored on film, but modern imaging modalities make it possible to turn the traditional film-based radiology into a filmless operation. All medical images today can be originally generated in digital form and stored in large medical Picture Archive and Communication Systems (PACS). Medical images are not just two-dimensional objects, but also offer three dimensions (e.g., volumetric or temporal 2D) and four dimensions (e.g., volumetric and temporal). Medical images are an important part of the EHR world and present some of the largest storage and management challenges to the electronic industry. Fortunately, all new medical imaging modalities and PACS work in conformance with the Digital Imaging and Communication in Medicine (DICOM) 3.0 standard [1] which has become both successful and complex. DICOM's success can be measured by its overwhelming and wide acceptance. This is the standard that unquestionably made filmless radiology a reality. DICOM is complex as it tries to cover more and more ground and has been evolving over the last decade.

The DICOM standard was created to resolve the challenge of enabling communication among all radiology imaging devices in order to facilitate their seamless integration into filmless radiology. DICOM is useful in operating the radiology department and integrating it within the hospital operation in general. DICOM is also crucial for telemedicine and for teleradiology in particular, where medical imaging is exchanged between remote locations for consultations, second opinions, and so forth. There is a tremendous amount of information in the DICOM format of medical images. Beyond the binary images, there are many attributes for the different modalities that describe imaging parameters and patient conditions (e.g., imaging-specific treatments and preparations [such as the injection of contrast agents] that the patient has gone through). DICOM also provides space to include manufacturer-specific attributes for backward compatibility and smooth transition of products from old protocols to DICOM.

DICOM defines two basic themes: data models of real world objects representing medical imaging exams, and communication protocols for exchange and inquiry of these objects. While many will argue that this is an oversimplification of the standard, it represents the majority of its existing applications. To complete and facilitate the management of a full hospital enterprise, DICOM has been combined with HL7 [2] to fit within a new whole standard called IHE (Integrating the Health Enterprise) [3]. For the scope of this paper, we take these two major themes of DICOM as the basis for the representation of and access to medical imaging as electronic health records.

Because EHR systems access many different data sources – including diverse modalities, PACS, HIS, CIS, RIS, and proprietary systems – they must comply with industry standards. We predict that the data in the EHR system will be represented in XML [4], as it is rapidly becoming the standard for data exchange, both across the web and among applications. Three main organizations are involved in creating standards related to EHR – HL7, CEN TC 215 [5] and ASTM E31 [6]. There are intentions to collaborate among the groups and to move toward the development of technically identical and interchangeable standards. This paper focuses on using the XML-based HL7 Clinical Document Architecture (CDA) standard [7], which provides an exchange model for clinical documents and brings the healthcare industry closer to the realization of an EHR. There are three different levels of CDA documents: CDA Level 1 represents a general clinical document, CDA Level 2 and Level 3 represent levels of specialization. At the time that this paper was written, only Level 1 was approved, while Level 2 and Level 3 were still in the initial development phase.

This paper describes how data stored in DICOM objects is transformed to XML based on the CDA standard. The EHR system can then index the XML documents by their markup as well as by their contents (much like the contents on the web). Complex queries can now be posed on the EHR system, such as, "Get all chest images of patients at the age of 20-30, that have blood type 'A' and are allergic to pine trees." A single PACS cannot answer such a query since patients may have images in multiple distinct facilities, and the execution of this query requires cross-checking with additional sources. In addition, by fully indexing all the data in DICOM objects, it becomes possible to offer access to huge amount of valuable data, which can be better exploited for research and treatment purposes in the specific radiology domain.

In the ensuing chapters of this paper we make some observations about the present state of issues relating to electronic health records. We set the stage to state a goal in which PACS fit into the broader scope of EHR. Finally, we present a solution for reaching this goal .

2. OBSERVATIONS

2.1 Diversity of data

Today, patient records are scattered among different archives both within and across institutions (nursing facilities, pharmacies, hospitals, physician offices). Not all are electronic and many of the computerized archives are not standardized, rather they are often based on proprietary and one-of-a-kind software. Moreover, the information in these archives is expressed in different vocabularies and terminologies, and in different formats and languages. When retrieved, the information is accessed and delivered using a variety of different and incompatible methods. As a result, in many cases the data that is stored at the point of origin is not available at point-of-care, unless these points are at the same location or closely related, as in a single hospital visit, a specific radiology procedure, or a procedure step. Medical patient history is stored and accessible somewhere, but is not always available and used when needed.

2.2 Medical research

Medical research in general is based on facts and as such, patient medical records are invaluable. In general, when specific experiments and investigations are conducted, special means are installed to collect the relevant medical data for the specific purpose of the research at hand in a controlled way. Sometimes, it is discovered that certain critical data that could be useful for the research has not been collected, which prevents the timely or satisfactory completion of important research topics. The question we raise is whether a complete collection of medical data could help alleviate such problems, or even prevent them. Clearly, the conditions under which the research needs to be conducted cannot in general be foreseen. But we suggest that when all the collected information is stored and fully indexed, the research community has the opportunity to investigate and analyze the vast amount of existing clinical data. Research can find correlations among various clinical attributes and facts whose value could not be foreseen at the moment of generation, collection, and storage.

2.3 Technologies that can help

The last few years have generated a vast and rich number of technologies for information processing, communications, knowledge management, integration of different applications into new ones, and more. The internet and its associated technologies (such as the web) are the cradle on which many new tools and standards flourish, creating numerous opportunities to quickly turn out new solutions. One of the fastest growing relevant standards is XML, which creates a common semi-structured data format that is both machine process-able, and human readable. Examples of these emerging technologies are tools that converge data from different XML sources into a combined unified representation using the XSL standard [8], tools to create extensive indexing of XML files – XMLFS [9], new storage solutions such as Network Attached Storage (NAS) [10], grid computing, which creates enormous computing power based on a large widely distributed collection of strongly connected computers [11]. On the other hand, there is the vast amount of computerized databases and repositories on the internet that provide data that can be used to

complement the patient's medical records, such as research articles, drug catalogues, and genomic databases, to name only a few.

3. GOALS

Our grand goal is to integrate all types of medical data into a single, extensively indexed collection. This collection will form the basis for patient electronic health records. An extensively indexed collection will allow the query, analysis, and correlation of data based on as full a representation as possible. Clearly, we require this collection to be bound to international standards and to use the most modern and powerful available tools. **Figure 1** depicts the position of this EHR in an information universe composed of sources and users. Sources are the plurality and divergent repositories of medical information, including patient health information, and other external information sources relevant to medical care. Users are both the human end users, such as doctors, care takers and patients, and other applications that use the EHR. The human end users access the information via GUI applications that understand the computer interface to EHR and that may relay this information in many different ways. They should not be limited in any way by the EHR machine interface. Therefore, we also include in this group pervasive, wireless, mobile devices as well as workstations and web clients. In addition, we envision computational resources that may perform further ongoing analysis of the data and generate new facts and findings. In essence, such applications may become new sources of data in this universe of medical information. In this figure, the grid is used to represent powerful analysis tools, such as those used for the nation-wide NDMA project [12] implemented in the United States.

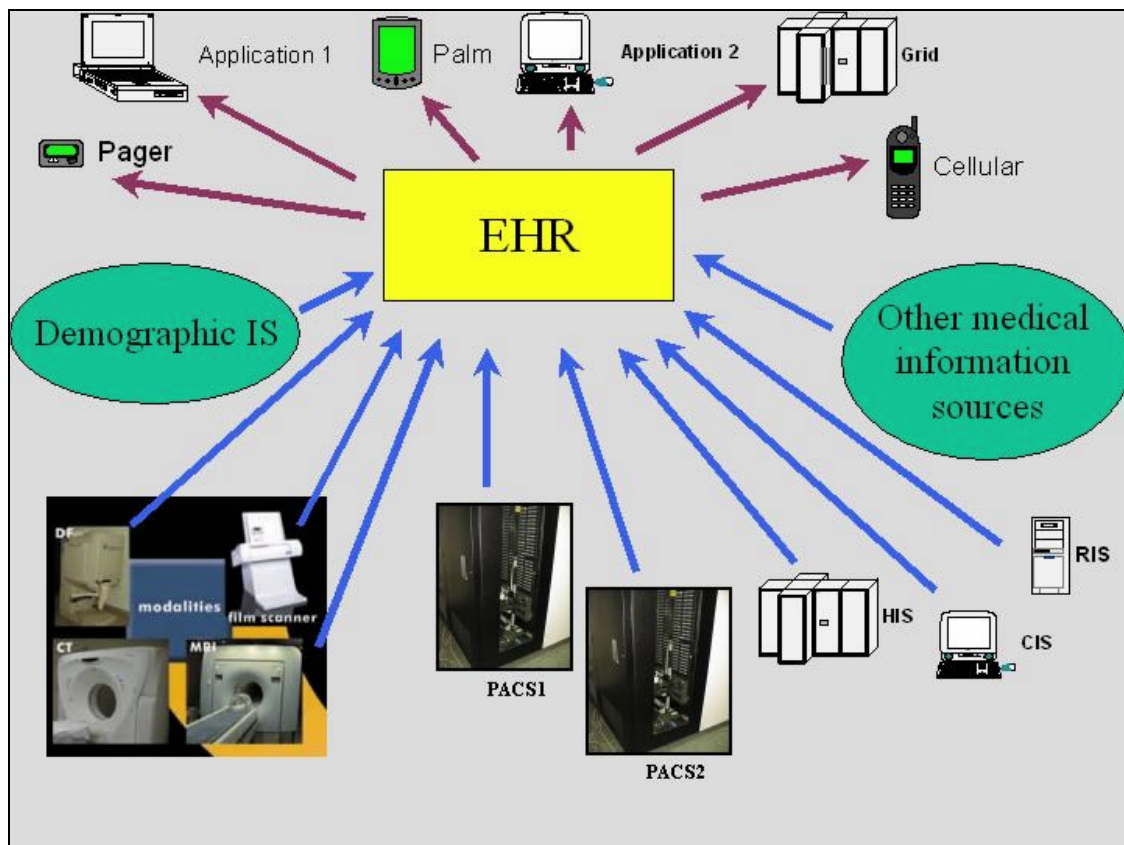


Figure 1: EHR System

Our initial stage is to integrate PACS information into the EHR. As PACS use the standard data model and communication protocol DICOM, this seems to be the most appropriate means. However, we believe that the format of choice for the electronic health record is XML, and indeed XML has been adopted by the HL7 version 3 standard. DICOM 3.0 is not XML based and, while it successfully fulfills its goal to move radiology towards becoming filmless (as described in the introduction above), we found that it lacks other important capabilities for EHR (described in section 4.1 below). Thus, the first step towards integrating PACS into EHR is to find a way to translate DICOM data to XML, based on the HL7 CDA standard.

4. INTEGRATING THE PACS INTO EHR

The major technical problem in converging image data from PACS into EHR is the conversion of the DICOM model, as projected through the DICOM communication protocols, into XML. More specific problem is the conversion of this information into legal CDA documents (which is XML based). In the following sections we will try to answer some of the intuitive questions raised by this plan, starting with the DICOM standard's weaknesses vis-a-vis our goals.

4.1 Is DICOM query sufficient?

The DICOM C-FIND and C-MOVE protocols perform Query/Retrieve services. It allows a client to find out, via a query, what images are stored in an archive or on another device, such as a workstation. Following that, the client may selectively retrieve a collection (such as all images of an exam) or individual images [13]. This query mechanism is quite limited, as stated in the standard documentation: "The types of queries, which are allowed, are not complex." (see [1], PS3.4, Annex C). The standard DICOM query can be used with three different information models, namely *Patient Root*, *Study Root*, and *Patient Study Only Information Models*. These possibilities are based on the hierarchical DICOM information model, which describes the real world as a hierarchical structure, composed of *Patient* entities that can have multiple *Studies*, each of which may hold several *Series*, each containing one or more *Images* (and/or other possible composite objects).

In the DICOM information model the "root" level of a query defines where to start the search in the tree hierarchy (Patient, Study, Series, Image). In each of the information models, only a limited number of key attributes can be used in the query. The keys in this set are categorized as either unique key attributes (U), required key attributes (R), or optional (O). Each level has a single unique search key, a few required keys, and the rest are optional. This set of query keys represents some pragmatic lowest common denominator needed to manage and run filmless radiology. The set of attributes in each level is not at all exhaustive. Some clinically important attributes are missing, such as *Body Part Examined* (0x0018, 0x0015), which has a significant value to EHR. Some justification for this fact may be that the DICOM query is actually used as a preliminary step to the retrieval of images (or other composite objects) and not as query for information only.

DICOM has defined an open standard in which comprehensive details and facts can be tagged in each image, series, and studies. More than that, private tags – as many as each vendor chooses – can also be added and stored. Of those tags whose type and internal formats are defined in the standard, the majority are usually optional and may be absent. While this approach is constructive in getting the standard widely accepted, it creates an information fabric that is full of holes. Nevertheless, in practice most of the attribute tags are filled in. Sometimes this depends on the good will and compliance of modality technicians, but the information in many cases flows into the PACS and is stored. However, even when stored, the query protocol is so limited that full use of the data requires it to be retrieved for further analysis. For example, if an orthopedic doctor wanted to track the changes in Jon Does' leg since its injury in a car accident, he would like to receive an answer to a query such as "provide all leg images of patient Jon Doe from the last 10 years." It seems that a DICOM query could not answer this. Therefore we conclude that the DICOM query is insufficient for EHR and that a stronger tool is required to fulfill this need.

4.2 Benefits of converting DICOM objects to XML

Transforming DICOM objects (without the binary pixel data and the private tags) to XML has important benefits, as demonstrated by the following examples. An open issue that still needs to be addressed is whether DICOM object representation in XML would be considered a CDA or not. The CDA defines documents as potentially authenticated. DICOM objects are observations, which are not authenticated per se. Each study and/or series may (and should) be associated with a report that can be authenticated, but these reports are not yet structured. Once the DICOM Structured Report (SR) [14] is harmonized with CDA Level 3 [15], correlation between the report and the image data will be made possible by this unified XML format.

The following examples show the power of search in data collected from PACS and transformed to indexed XML documents. To answer those queries in the original PACS system, one would need an exhaustive retrieve of all stored DICOM objects.

Example 1: Medical research applications, using information stored in different PACS

The obvious benefit of integrating different medical data and allowing search on this data, is to supply research applications with a tool that can pose complex queries across all patients, such as, "Get all CT chest images of male patients, age 20-30, living in Canada, working in petrochemical industry, with white cell count above normal."

Specifically, in the radiology domain, the transformation of DICOM objects to XML and indexing of all the data enables such complex queries. Or another example, "Get all CT chest images of men, age of 30-45, weighting above 100 kg." Note that all of this information is present in DICOM objects, but cannot be specified directly in the DICOM query since the attribute *Body Part Examined* is not a key attribute.

Example 2: Planning contrast agents administration procedures

Today, doctors decide on the proper dose of contrast agents and the timing of the procedure according to some pre-calculated percentage depending on weight, age (in pediatric departments), manufacturer's recommendation, and personal experience. One of the modules in the Image Information Entity of CT, MRI and other modalities (if contrast agents are used) includes contrast attributes, such as *Contrast Bolus Agent*, *Contrast Bolus Agent Sequence*, *Contrast Bolus Volume*, as well as the *Start Time*, *Stop Time*, *Total Dose*, *Flow Rate*, *Flow Duration*, and so on. When DICOM objects are transformed to indexed XML documents, queries can be performed based on these tags. This enables research on planning and optimizing a procedure for administering contrast agents. In addition, the procedure may be personalized by correlating information about the patient's medical condition. For example, a patient with heart failure condition (CHF) might need a higher dose of contrast agent for the image to be of sufficient quality. The benefits are obvious: administering the correct dose eliminates the need for repeat tests due to incorrect dosage. This saves costs, both in terms of conducting fewer tests and using contrast agent material.

Example 3: Statistical analysis

Statistical analysis can be done to enhance control (including budget control) on different modalities or multiple devices of the same modality. This can be of special interest for large hospitals. A decision such as choosing a modality manufacturer or PACS vendor can be based on such analysis. The following episodes demonstrate this use of statistical analysis: (1) Tracking the number of CT and MRI images taken on different body parts using various modalities, in order to answer questions such as "Which modality model is better for examining a specific body part?" (2) Tracking the number of CR, US, CT, MRI images taken per day in order to optimize scheduling issues or assign technicians to different modalities. (3) A hospital plans to buy a new modality can use information gathered in its own facilities to differentiate the quality, throughput, and operational costs of relevant in-house modalities, which can help in the purchase decision.

Example 4: Planning physical procedure in the image acquisition

A DICOM image holds a great deal of technical information that can be extracted by queries on EHRs and analyzed in order to achieve better image quality and resolution, and to plan physical protocols. For example, a CT image contains attributes such as *X-Ray Tube Current*, *Distance Source To Detector*, *Distance Source To Patient*, and others. MRI images contain attributes like *Repetition Time*, *Echo Time*, *Magnetic Field Strength*, and more. These attributes influence image quality. Research can be done on the existing data, in order to improve the acquisition of the images, by posing a query such as, “Get me all MRI images obtained from the modality model SignaSP manufactured by GE” and correlate it with the specific attribute under investigation.

4.3 DICOM to XML conversion

The XML format of the DICOM object is based on the CDA standard and includes two main elements: the *clinical_document_header* element and the *body* element. The *clinical_document_header* element is composed of four logical components: document information, encounter data, service providers, and patient information. The values suitable for the *clinical_document_header* subtree are extracted from the DICOM object relevant tags. This process requires mapping between DICOM data types and HL7 data types (see Appendix), as well as matching codes from the HL7 RIM and vocabulary.

The *body* element represents a DICOM image, as defined by the standard. For example, the CT image Information Object Definition (IOD) defines the content and structure of a CT image. It defines which DICOM attributes are mandatory, which are optional, and the multiplicity of each attribute. **Table 1**, which is taken from the DICOM standard, describes which modules will be included in the DICOM image object. The *body* subtree of a CT image is constructed accordingly, and each module holds certain attributes as designated by DICOM.

IE	Module	Reference	Usage
Patient	Patient	C.7.1.1	M
Study	General Study	C.7.2.1	M
	Patient Study	C.7.2.2	U
Series	General Series	C.7.3.1	M
Frame of Reference	Frame of Reference	C.7.4.1	M
Equipment	General Equipment	C.7.5.1	M
Image	General Image	C.7.6.1	M
	Image Plane	C.7.6.2	M
	Image Pixel	C.7.6.3	M
	Contrast/bolus	C.7.6.4	C - Required if contrast media was used in this image
	CT Image	C.8.2.1	M
	Overlay Plane	C.9.2	U
	VOI LUT	C.11.2	U
	SOP Common	C.12.1	M

Table 1: CT Image IOD (Information Object Definition) Module Table [1, PS3, A3.3]

The following frame represents part of the DTD for the *body* element:

```
<!ELEMENT body (patient_information_entity, study_information_entity,
series_information_entity, frame_of_reference_information_entity,
general_equipment_information_entity, image_information_entity+ )>

<!ELEMENT patient_information_entity (patient_module)>
<!ELEMENT study_information_entity (general_study_module,
patient_study_module?)>
<!ELEMENT series_information_entity (general_series_module)>
<!ELEMENT frame_of_reference_information_entity (frame_of_reference_module)>
<!ELEMENT general_equipment_information_entity (general_equipment_module)>
<!ELEMENT image_information_entity ( general_image_module,
image_plane_module, image_pixel_module, contrast_bolus_module?,
ct_image_module, overlay_plane_module?, voi_lut_module?, sop_common_module
)>
```

Each module in the DTD contains DICOM attributes. For example, DICOM attributes for the Patient Study Module are described in **Table 2**.

Attribute Name	Tag	Type	Attribute Description
Admitting Diagnoses Description	(0008,1080)	3	Description of the admitting diagnosis (diagnoses)
Patient's Age	(0010,1010)	3	Age of the Patient.
Patient's Size	(0010,1020)	3	Length or size of the Patient, in meters.
Patient's Weight	(0010,1030)	3	Weight of the Patient, in kilograms.
Occupation	(0010,2180)	3	Occupation of the Patient.
Additional Patient's History	(0010,21B0)	3	Additional information about the Patient's medical history.

Table 2: Patient Study Module Attributes

The respective DTD for the *patient_study_module* element is:

```
<!ELEMENT patient_study_module (admitting_diagnosis_description?,
patient_age?, patient_size?, patient_weight?, occupation?,
additional_patient_history? )>
```

According to the DICOM Data Dictionary (see [1], PS3.6), each DICOM element holds a numeric tag, consisting of group and element hexadecimal numbers. This is described in the DTD by two fixed attributes for each DICOM element. These attributes are the group and element numbers. In our example, *Patient Age*, (0x0010,0x1010).

This part of the DTD looks as follows:

```
<!ENTITY    %dicom_AS    "(#PCDATA)">
<!ELEMENT  patient_age    %dicom_AS;>
  ATTLIST  patient_age
    group  CDATA #FIXED "0010
    element CDATA #FIXED "1010">
```

In the XML document, each DICOM attribute, which is an XML element, has the two attributes described above and a value extracted from the DICOM object. To continue our example, the *Patient Age* will appear as follows:

```
<patient_study_module>
  ....
  <patient_age  group="0010"  element="1010">
    031Y
  </patient_age>
</ patient_study_module>
```

An exception to this transformation method is the binary-valued DICOM attribute *Pixel Data* (0x7FE0,0x0010). In this case, the XML document will not include the binary image data, but will provide information to allow retrieval of this data when and if needed from its original repository.

The full proposal to convert DICOM 3.0 to CDA Level 3 was sent to the HL7 Structured Document Technical Committee, and can be found in [16]. IBM research project SHAMAN/IMR [17] includes an initial implementation of this conversion.

5. SUMMARY

There is a significant amount of medical data available out there, and it is diverse and, for the most part, not standardized. We believe that the combined whole is larger than the sum of its parts, and envision that medicine of the future will include EHR systems. These systems will collect and convert data from dispersed sources to XML. This view is strengthened by the fact that international organizations are already developing XML-based standards for the healthcare domain, such as HL7 CDA. The EHR system will include an extensive indexing and query mechanism, much like the web. Various organizations such as care providers, research institutes, pharmaceutical companies, FDA, and so on, will use EHR systems to perform data mining, analyze and correlate the vast amount of medical data.

An initial step in creating such EHR systems is converting the medical images scattered in various PACS to XML documents. This paper proposes a method for such conversion as well as describes in general terms various applications that explore the medical images in a way that could not be done by DICOM alone. For example, we suggest applications to improve contrast agent administration and radiation dose procedures based on analyzing the medical images data in the EHR system.

Some of the ideas presented here were implemented in an IBM research project done in collaboration with hospitals in Israel, specifically with the radiology department of the Tel Hashomer Sheba hospital near Tel-Aviv, Israel.

ACKNOWLEDGEMENTS

Thanks to Pnina Vortman, Yevgeny Burshtein, and Amnon Shabo of the Multimedia and CRM group of the IBM Haifa Research Lab who participated and contributed to the work reported here. Nancy Ozeri who helped with editing, and to Dr. Jakobson of the Tel-Hashomer Sheba Hospital Radiology Department for his kind advice.

REFERENCES

1. Digital Imaging and Communication in Medicine (DICOM), National Electrical Manufacturers Association, <http://medical.nema.org/dicom/2001.html>
2. Health Level 7 (HL7), <http://www.hl7.org/>
3. Integrating the Health Enterprise (IHE), <http://www.rsna.org/IHE/index.shtml>
4. eXtensible Markup Language (XML), <http://www.w3c.org/XML>
5. CEN TC 251, <http://www.cen251.org/>
6. ASTM E31, <http://www.astm.org/cgi-bin/SoftCart.exe/COMMIT/COMMITTEE/E31.htm?L+mystore+txf8000+970175387>
7. Clinical Document Architecture Framework, Release 1.0, Liora Alschuler, Robert H. Dolin, Sandy Boyer, Calvin Beebe, Paul V. Biron, Rachael Sokolowski.
8. "Extensible Stylesheet Language (XSL) Version 1.0," W3C Recommendations, 15 October, 2001 (<http://www.w3.org/TR/xsl/>)
9. Carmel, D., Maarek, Y, and Sofer, A. "XML and Information Retrieval, a SIGIR 2000 Workshop". In SIGIR Forum, 34:1, 2000 (<http://www.haifa.il.ibm.com/projects/systems/xmlfs/xmlfs-jasis.pdf>).
10. David F. Nagle et. al., "Network Support for Network-Attached Storage," Proceedings of Hot Interconnects 1999, August 18-20, 1999, Stanford University, CA (<http://www.pdl.cs.cmu.edu/PDL-FTP/NASD/hotnet99.pdf>)
11. Oster, I. and Kesselman, C. (eds.), "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann, 1999.
12. "NDMA - National Digital Mammography Archive," (<http://nscp01.physics.upenn.edu/ndma/projovw.htm>)
13. Oosterwijk, H. "DICOM basics", 2000.
14. DICOM SR, supplement 23: "Structured Reporting," NEMA, Rosslyn VA, 2000, ftp://medical.nema.org/MEDICAL/Dicom/Final/sup23_ft.pdf
15. Behelen, F. "Using XML tools in DICOM."
16. CDA Level 3 for DICOM Series, http://www.haifa.il.ibm.com/projects/software/imr/papers/DICOM_CDA_ver0.2.zip
17. SHAMAN/ IMR project, <http://www.haifa.il.ibm.com/projects/software/imr/index.html>

APPENDIX

Mapping DICOM Data Types to HL7 Data Types – an initial proposal

While mapping specific DICOM data types to the more general HL7 data types, some information is lost. In addition, DICOM defines Value Representation (VR) for data elements. VR describes the data type and format of the data element value. For example, the DICOM data type AS (for Age String) is defined as: “A string of characters with one of the following formats -- nnnD, nnnW, nnnM, nnnY; where nnn shall contain the number of days for D, weeks for W, months for M, or years for Y. For example, ‘018M’ would represent an age of 18 months.” Mapping it to HL7 type ST (for Character String), loses the implicit information of the original data format.

Note: To improve compliance with DICOM data types, it would be better to use an XML Schema instead of a DTD. This has not been done yet.

Table 3: Mapping DICOM VRs to HL7 Data Types, based on DICOM and the HL7 Data Types Part I (<http://www.hl7.org/home/content/HTML-menus/DataTypes1.htm> - TS).

DICOM Data Types		HL7 Data Types		Comments
VR *	Full Name	Symbol	Full Name	
AE	Application Entity	ST	Character String	
AS	Age String	ST	Character String	
AT	Attribute Tag **	SET	Set	Need a conversion function
CS	Code String	ST	Character String	
DA	Date	TS	Point in Time	By omitting the hhmmss.uuuu part
DS	Decimal String	REAL	Real Number	Need a conversion function
DT	Date Time	TS	Point in Time	
FL	Floating Point Single	REAL	Real Number	
FD	Floating Point Double	REAL	Real Number	
IS	Integer String	INT	Integer Number	Need a conversion function
LO	Long String	ST	Character String	
LT	Long Text	ST	Character String	
OB	Other Byte String **	ST	Character String	Not good enough
OW	Other Word String **	ST	Character String	Not good enough
PN	Person Name	PN	Person Name	
SH	Short String	ST	Character String	
SL	Signed Long	INT	Integer Number	
SQ	Sequence of Items	LIST	Sequence	
SS	Signed Short	INT	Integer Number	
ST	Short Text	ST	Character String	
TM	Time	?		
UI	Unique Identifier (UID)	II	Instance Identifier	
UL	Unsigned Long	INT	Integer Number	
UN	Unknown **	ST	Character String	Not good enough
US	Unsigned Short	INT	Integer Number	
UT	Unlimited Text	ST	Character String	