Using the XML-based Clinical Document Architecture for Exchange of Structured Discharge Summaries

Grace I. Paterson[a], Michael Shepherd[b], Xiaoli Wang[b], Carolyn Watters[b], David Zitner[a]
[a]*Medical Informatics, Faculty of Medicine,*
[b]*Faculty of Computer Science*
Dalhousie University
Halifax, Nova Scotia, Canada
{grace.paterson | david.zitner}} @ dal.ca
{shepherd | watters | xiaoli} @ cs.dal.ca

Abstract

The XML-based Clinical Document Architecture (CDA) for document exchange defines a three-level document architecture with each higher level adding more specificity to the markup of the document. Such a layered architecture also meets the requirement for semantic processing of hierarchically structured clinical documents. The CDA supports shared care between hospital-based and community-based physicians, knowledge integration by permitting external links to other documents, and outcomes research through the capture of discrete and coded clinical data.

The research reported in this paper has lead to the successful prototyping of a CDA-based Structured Discharge Summary system. Structured discharge summaries are used to move clinical information from the clinical environment, such as the geriatrics department of a hospital, to the community environment of family practice. They are an education, communication, research, and administration tool in the Departments of Medicine and Family Medicine and at the Queen Elizabeth II Health Sciences Centre.

1. Introduction

With the rapid development of the Internet and the World Wide Web, the computer-to-computer exchange of business documents in a structured, predefined standard format has become more and more significant. This exchange has been greatly facilitated by the development of extensible structured markup languages such as XML. A similar need for the electronic exchange of clinical documents exists in healthcare. Until recently, however, standards for clinical document exchange among clinical systems covered messaging of fielded data but did not meet the need for semantic processing of hierarchical, structured, clinical documents. This has now been addressed by the Clinical Document Architecture (CDA).

ANSI/HL7 CDA R1.0-2000 is the first nationally certified XML-based standard for healthcare [1]. It has been developed by Health Level 7 (HL7), an ANSI-accredited Standards Developing Organization operating in the healthcare arena. HL7's mission is to enable clinical interoperability through the provision of "… standards for the exchange, management and integration of data that support clinical patient care and the management, delivery and evaluation of healthcare services." [10].

The Clinical Document Architecture has been in development since 1996, originally as the Kona Architecture [11], then as the Patient Record Architecture (PRA), and now as the CDA [5]. It is a document markup standard for the structure and semantics of exchanged "clinical documents". A CDA document is a defined and complete information object that can exist outside of a message and can include text, images, sounds, and other multimedia content.

The CDA is a three-level DTD with each higher level adding more specificity to the markup of the document. At the time of writing, only the Level One DTD has been fully defined. Level One is the root of the hierarchy and is the most general DTD. Although it is the most general, it is possible to differentiate types of documents, such as a "Referral Form" from a "Discharge Summary", because these are different document type code values in the document instance. Above Level One, it will be possible to specify additional constraints on a document by creating distinct XML DTDs for each type of document, such as Cardiology Discharge Summary or Respirology Discharge Summary. The clinical content will not vary across levels, just the degree to which clinical content can be specified and processed. This allows flexibility

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

depending on the requirements of the application. Thus, a clinical document marked up according to a Level Three DTD can still be processed by an application requiring only the Level One markup.

Levels are a migration pathway for adding more markup. Levels Two and Three are still under development. Level One uses a document header based on the semantics in the open information data model, HL7 Reference Information Model (RIM), but the body of the document is largely unstructured. This minimizes the technical barriers to entry, while starting users on a migration path to use RIM. Level Two introduces more iterations of Level One, but it is expected that Level Three will be available next. Clinical content in Level Three can be marked up to the extent that it is modeled in the RIM [7]. The dissemination of Level Three specification awaits tabulation of the international ballot on Version 3 of RIM. There are six distinct ballot groups in different stages of committee and membership ballot, and tabulation is not anticipated until 2002.

The prototype application developed in this research was designed using CDA Level One to handle Web-based data entry and retrieval of structured discharge summaries. Discharge summaries are used to move clinical information from the clinical environment, such as the cardiology department of a hospital, to the community environment of family practice. In the prototype, the structured discharge summaries are document instances of the XML-based Clinical Document Architecture (CDA). A structured discharge summary follows a template with regard to content and form, and information can be added to improve the quality of the information, such as the outcomes of care. The user interface is a Web browser, for both viewing and editing the discharge summary, and a native XML database system is used for storing the XML documents.

Figure 1 shows the home page with access to the two document types available, Discharge Summary and Referral Form.

The Web-form discharge summary is a contiguous form presented in a scrollable window in a Web browser. Figures 2 and 3 are the first two parts of such a form; the rest of the form has not been included in this paper. Such a form can be used for either display or editing. As information is added to the discharge summary at various points in the medical system, this information is captured, marked-up in XML, and stored on the server. This discharge summary can then be accessed at the next point in the medical system and viewed, edited and stored back on the server. The XML-based summary may itself by moved to another server and interpreted at that server by CDA-aware applications. Figure 4 is the same information presented in a more standard document format within the Web browser.

The system is implemented by a three-tier Web architecture, using XML, HTML, JavaScript, SAX and Java Servlets. The structured discharge summary forms are patient record documents that serve as an education, communication, research, and administration tool in the Departments of Medicine and Family Medicine, Queen Elizabeth II Health Sciences Centre, Halifax, NS, Canada.

Section 2 of this paper briefly presents some background on how healthcare information is exchanged electronically while Section 3 presents an overview of the Clinical Document Architecture with examples from the Structured Discharge Summaries used in this research. Section 4 discusses the three-tier architecture of the system and Section 5 summarizes this work.

2. Exchange of healthcare information

The Standard Generalized Markup Language (SGML) is a "meta-language" that supports the document-centered approach to medical records [16]. XML or the eXtensible Markup Language is an application of SGML for digitally stored documents that contain structured information. This information can be in the form of text, pictures, and biological signals.

There are non-CDA implementations of XML for healthcare. The first United Kingdom project to send and receive discharge letters using XML as a message format was implemented in Scotland [19] and integrated discharge communication with the nationally distributed General Practice Administration System for Scotland (GPASS) [20]. Java and XML have also been used in telemedicine for the secure distribution of medical records [9]. XML is a human and machine readable language so lends itself to a wide range of healthcare communication needs.

For the health care sector, the two standards for message exchange are distinguished as within or outside the organization. EDIFACT messages, which are generally used for communication with systems outside the own organization, and the HL7 standard, which is generally used for communication within an organization [3]. The European MEDSEC project addresses communication security in open health systems, and the need to enable inter-organizational data exchange, such as sharing care between a developed and developing country. This work provided content for the HL7 Security Service Framework and demonstrated use of XML as a standardized communications technique [4]. With the emergence of the XML-based CDA, electronic communication between the hospital-based specialist and the community-based family physicians has been facilitated. This XML-based CDA is a standard approved by Canada HL7 , and was presented at the HL7 International Affiliates Joint Meeting in Germany [6].

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

Figure 1. Opening screen

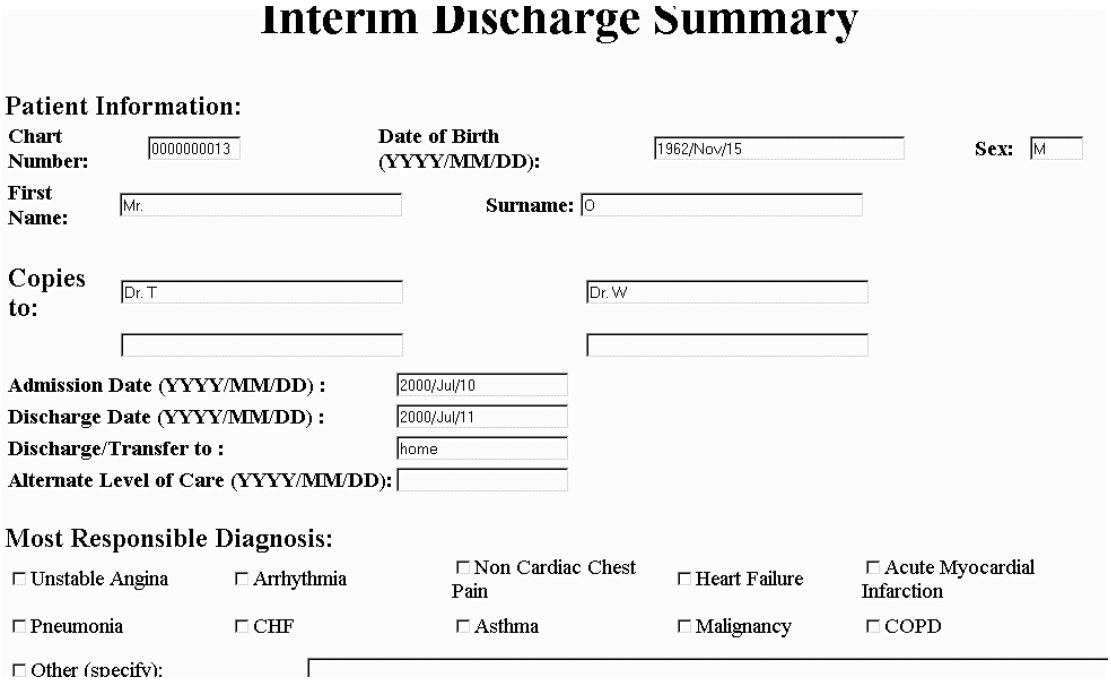
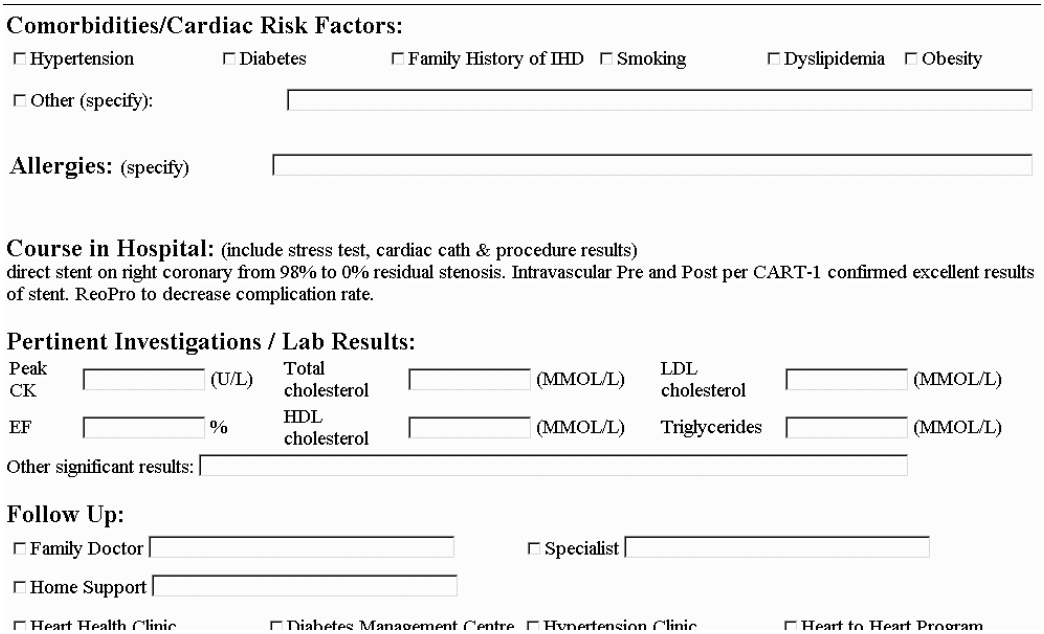


Figure 2. First part of Web-form display of discharge summary



Comorbidities/Cardiac Risk Factors:

☐ Hypertension ☐ Diabetes ☐ Family History of IHD ☐ Smoking ☐ Dyslipidemia ☐ Obesity

☐ Other (specify):

Allergies: (specify)

Course in Hospital: (include stress test, cardiac cath & procedure results)
direct stent on right coronary from 98% to 0% residual stenosis. Intravascular Pre and Post per CART-1 confirmed excellent results of stent. ReoPro to decrease complication rate.

Pertinent Investigations / Lab Results:

Peak CK	(U/L)	Total cholesterol	(MMOL/L)	LDL cholesterol	(MMOL/L)
EF	%	HDL cholesterol	(MMOL/L)	Triglycerides	(MMOL/L)

Other significant results:

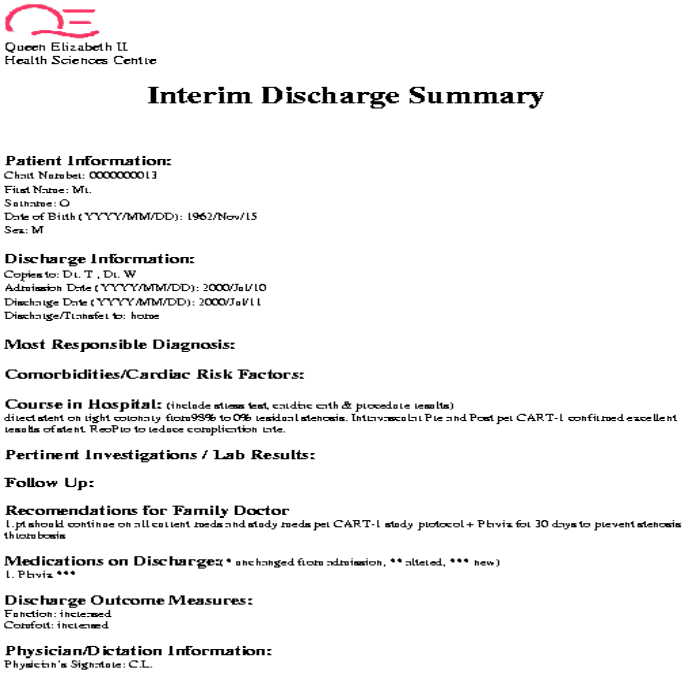
Follow Up:

☐ Family Doctor ☐ Specialist

☐ Home Support

☐ Heart Health Clinic ☐ Diabetes Management Centre ☐ Hypertension Clinic ☐ Heart to Heart Program

Figure 3. Second part of Web-form display of discharge summary



Queen Elizabeth II
Health Sciences Centre

Interim Discharge Summary

Patient Information:
Chart Number: 0000000013
First Name: Mt.
Surname: O
Date of Birth (YYYY/MM/DD): 1962/Nov/15
Sex: M

Discharge Information:
Copies to: Dr. T , Dr. W
Admission Date (YYYY/MM/DD): 2000/Jul/10
Discharge Date (YYYY/MM/DD): 2000/Jul/11
Discharge/Transfer to: home

Most Responsible Diagnosis:

Comorbidities/Cardiac Risk Factors:

Course in Hospital: (include stress test, cardiac cath & procedure results)
direct stent on right coronary from 98% to 0% residual stenosis. Intravascular Pre and Post per CART-1 confirmed excellent results of stent. ReoPro to reduce complication rate.

Pertinent Investigations / Lab Results:

Follow Up:

Recomendations for Family Doctor
1. pt should continue on all current meds and study meds per CART-1 study protocol + Plavix for 30 days to prevent stenosis thrombosis

Medications on Discharge: (* unchanged from admission, ** altered, *** new)
1. Plavix ***

Discharge Outcome Measures:
Function: increased
Comfort: increased

Physician/Dictation Information:
Physician's Signature: C.L.

Figure 4. Document form of discharge summary

Clinical terminology and the use of controlled vocabularies, such as SNOMED and the International Classification of Diseases revision 10 (ICD10), have been well studied and are vital to the exchange of healthcare information. A medical language processing system using XML to support vocabulary development has been used to process pathology and radiology reports [12]. The HL7 vocabulary domains have been reviewed to ensure they support not only direct clinical care but also clinical research, outcomes research, and population health management [2].

The intent of the ASTM subcommittee, ASTM E31.25, is to provide XML Document Type Definitions (DTDs) for Health Care. The focus is document-centric, as opposed to the document exchange focus of CDA [17].

3. CDA overview

CDA documents derive their meaning from the HL7 Reference Information Model (RIM). The RIM provides a coherent shared information model that contains all data content relevant to HL7 messages, and is an essential part of the HL7 Version 3 development methodology. It represents the semantic and lexical connections between the information carried in the fields of HL7 messages and has evolved to a flexible and general model of clinical information.

The Clinical Document Architecture is a three-layer architecture implemented in XML, where each level is defined by a DTD. Level One is the root of the hierarchy and each additional level adds further specificity and constraints to the architecture. Level One specifies the semantics of the header, codes for the document type and sections within the body of the document [8]. Level Two uses the same codes as Level One for the document type and sections but will allow further constraints to be imposed. Level Three will define observations and services within the document body. At this time, Levels Two and Three have not yet been fully defined by HL7.

Level One consists of three technical specifications; the CDA Header, the CDA Level One Body, and the HL7 Version 3 data types. Figure 5 illustrates the overall structure of a Level One Document Instance of a structured discharge summary in the prototype system.

3.1 HL7 Version 3 Data Types

As indicated above, CDA documents derive their meaning from the HL7 Reference Information Model (RIM). The elements and attributes and the relationships among these elements and attributes are drawn from the RIM and expressed in XML.

The HL7 Version 3 Data Types ballot is the specification that attempts to define all the data types needed for health care information exchange [14]. These data types provide semantic constraints for the attributes of the RIM classes. Data types are defined for (1) character strings and display data, which accommodates both character based text and multimedia data; (2) codes and identifiers for concepts and instances both of the real world and of technical artifacts; (3) all kinds of quantities including integer and real numbers, physical measurements with units, various kinds of time. Canada HL7 provides specifications in both English and French, so a document can be expressed in either language using Version 3 Data Types.

Coded CDA components or elements have the two-letter ending, “cd”. These coded components have associated vocabulary domains, which represent allowable value sets for the components. These domains may be HL7-defined concepts or recognized coding schemes such as SNOMED.

```
<?xml version="1.0"?>
<!DOCTYPE levelone PUBLIC "-//HL7//DTD CDA Level One 1.0//EN" "dischargesummary.dtd">
<levelone>
   <clinical_document_header>
…

   </ clinical_document_header>
   <body confidentiality= “CONF1”>
…

   </body>
</levelone>
```

Figure 5. Overall structure of Level One document instance

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

Every vocabulary has a unique HL7-assigned identifier, and every concept within a vocabulary domain has a unique code. The root HL7 Object Identifier is 2.16.840.1.113883, and the steward for the codes is the Vocabulary Committee. As an example, the definition for object identifier 2.16.840.1.113883.6.3 is the external coding scheme, International Classification of Diseases revision 10 (ICD10). These coded components are expressed as attributes of the XML elements.

For example, Figure 6 illustrates the use of the caption, coded caption and vocabulary domains in the Body of the document instance:

```
<caption>Most Responsible Diagnosis</caption>
<section>
   <caption>Unstable Angina
      <caption_cd V=I20.0 S="2.16.840.1.113883.6.3"/>
      </caption>
      <paragraph>
         <content>Y</content>
      </paragraph>
</section>
```

Figure 6. Captions, coded captions and vocabulary

The caption for the section is "Most Responsible Diagnosis". The caption for the paragraph is "Unstable Angina". The code representing Unstable Angina is I20.0 and is taken from the source vocabulary, ICD10. This permits the caption to be displayed and/or read by a human and the corresponding code to be processed by a computer application.

3.2 CDA Header

The CDA Header contains the metadata describing this clinical document. It consists of four logical components:

- Document information, including relationships to other documents
- Encounter data
- Service actors (such as providers)
- Service targets (such as patients)

The major elements of the clinical_document_header used in this prototype include:

- id
- set_id
- version_nbr
- document_type_cd
- origination_dttm
- confidentiality_cd
- document_relationship
- patient_encounter
- legal_authenticator
- originator
- originating_organization
- originating_device
- provider
- patient
- local_header

Many of these elements may have sub-elements. For instance, as illustrated in Figure 8, the patient element may have sub-elements and an example may be coded as in Figure 7.

```
<patient>
   <patient.type_cd V="PATSBJ"/>
   <person>
      <id EX="12345" RT="2.16.840.1.113883.3.933"/>
      <person_name>
         <nm>
            <GIV V="John"/>
            <FAM V="Doe"/>
         </nm>
         <person_name.type_cd V="L"
               S="2.16.840.1.113883.5.200"/>
      </person_name>
   </person>
   <birth_dttm V="1932-09-24"/>
   <administrative_gender_cd V="M"
      S="2.16.840.1.113883.5.1"/>
</patient>
```

Figure 7. Example of coding sub-elements

3.2 CDA Body

The CDA Level One body is comprised of nested containers, including non_xml data, sections, paragraphs, lists, and tables. The sections may be nested and may also contain paragraphs, lists and tables, as indicated in Figure 8. The containers may have captions and contents and may be coded. Each container also has confidentiality and origination attributes. This enables the sharing of information to groups with different confidentiality levels. Figure 9 is a partial example of a Level One Body coding from the prototype. It shows sections with captions and contents. The entire body has the confidentiality level of CONF1.

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

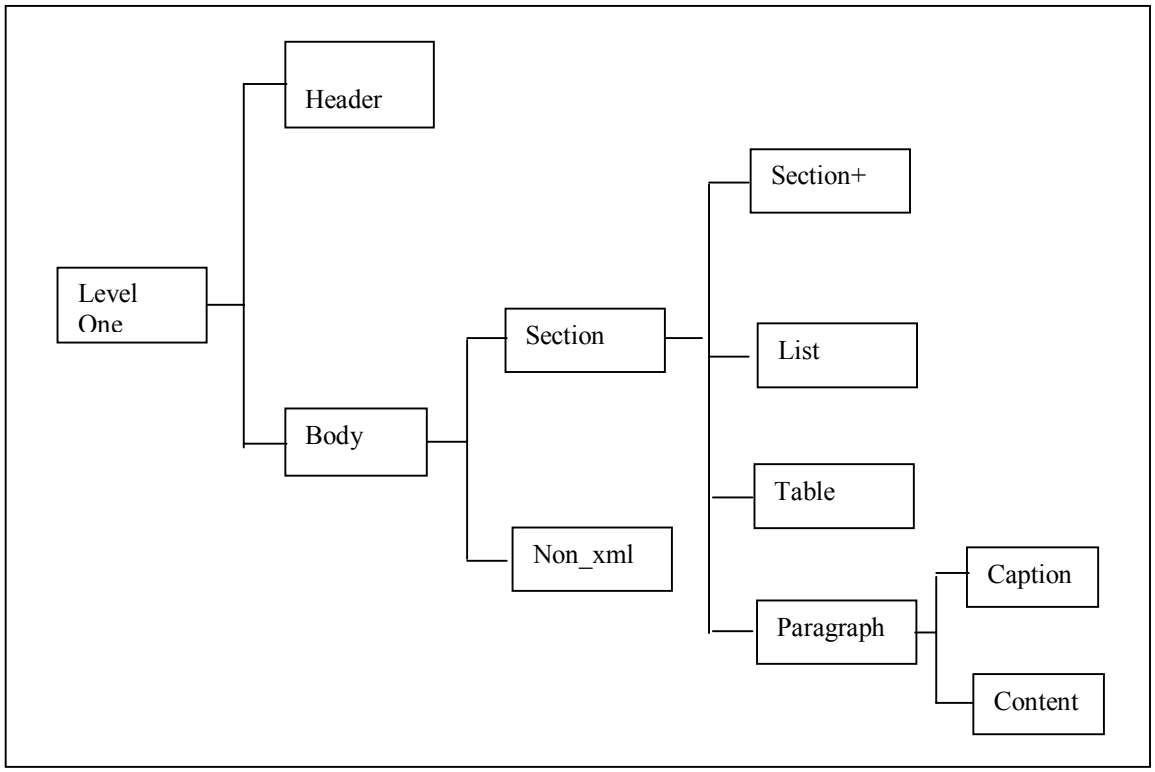


Figure 8. Hierarchical Container Structure of Level One.

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 **IEEE**

```
<body confidentiality= “CONF1”>
   <section>
      <caption>
         <caption_cd V="8684-3" S="2.16.840.1.113883.6.1"/>
      </caption>
   </section>
   <section>
      <caption>Discharge Summary</caption>
      </section>
   <section>
      <caption>Patient Information</caption>
      <section>
         <section>
            <caption>Chart Number</caption>
            <paragraph>
               <content>12</content>
            </paragraph>
         </section>
         <section>
            <caption>Date of Birth</caption>
            <paragraph>
               <content>1940/Jan/14</content>
            </paragraph>
         </section>
      </section>
         …
      <section>
         <caption>Most Responsible Diagnosis</caption>
         <section>
            <caption>Unstable Angina
               <caption_cd V=I20.0 S="2.16.840.1.113883.6.3"/>
            </caption>
            <paragraph>
               <content>Y</content>
            </paragraph>
         </section>
      …
</body>
```

Figure 9. Example of coded Level One Body.

4. Architecture

The architecture of the prototype system is a typical three-layered architecture. The presentation layer will run with any JavaScript enabled browser. JavaScript and HTML implement the interface and interact with the user as information is entered into the form.

The portal has three features:

1) The ability to add, edit, find, and query discharge summaries
2) The ability to add, edit, find, and query referral forms
3) The ability to list all discharges and referrals assigned to a given doctor

To generate a Discharge Summary or Referral Form, the physician enters the appropriate information at the Web browser. This information is POSTed to the Web server where a Java servlet parses the input data and

marks up the data into the CDA format. The servlet then inserts the record into the dbXML database.

Upon request for viewing a Discharge Summary, the user inputs the chart or referral number and the type of display. A Java servlet retrieves the selected record in the dbXML database. If it is found, the XML file is first parsed for validation using the corresponding SAXParser. The appropriate elements are then converted to HTML and sent to the browser for display. The editing is handled similarly, using a servlet, the SAXParser, and display for editing. Any changes will overwrite the existing record.

The query finds all patients with a particular diagnosis using radio buttons for a predefined set of diagnoses. The input is sent to a servlet. The servlet then executes an Xpath query into the dbXML database and returns a vector of XML records that satisfies the query.

The listing by doctor function is handled using a servlet that executes an Xpath query into the dbXML dictate collection. This method retrieves a vector of XML records where the doctor's name is found. The same servlet will also execute an Xpath query into the dbXML referral collection to retrieve a vector of XML records where the doctor's name is found. The query on diagnosis returns the set of record identifiers of all XML documents in the database with that diagnosis.

The demonstration system uses a native XML database, dbXML. Although there are other storage strategies for CDA Documents such as, XML-Enabled Database in a relational DBMS or an Object Database Management System [18], dbXML was chosen because of data complexity and performance as it was designed specifically for XML documents.

5. Summary

The transfer of patient information from a clinical environment to the family physician is a vital aspect of our healthcare system and is important in providing overall care for the patient. Family physicians are under constant time pressure, and improving the quality of the discharge summary and standardizing the elements of that summary will improve the nature of the care. Shared information that is pertinent to patient care saves time and resources. Shared knowledge around diagnostic and treatment recommendations provides decision-support for caregivers and patients. Standardizing the Structured Discharge Summary goes a long way to providing the required information and knowledge.

CDA for document exchange supports shared care between hospital-based and community-based physicians, knowledge integration by permitting external links to other documents, and outcomes research through capture of discrete and coded clinical data [13]. Health information that can be coded at time of capture, saves time and money for health system administration, and provides an enhanced information resource to serve needs of caregivers, researchers, and educators.

The XML-based Clinical Document Architecture is a flexible and powerful vehicle for the electronic exchange of such discharge summaries. However, the CDA goes far beyond just discharge summaries. It can be used for virtually any type of clinical document and can become part of the electronic patient record. As well, as it is Web-based, such clinical documents instances can be accessed through medical portals [15] on either a push and/or pull basis across the Web.

The prototype presented in this paper illustrates the use of the Clinical Document Architecture for the exchange of discharge summaries. Further work is proceeding to enrich the summaries themselves with more information and, eventually, proceed to Levels Two and Three of the CDA.

6. References

[1] L. Alshuler and R. Dolin. (Ed.) *Version 3 Standard: Clinical Document Architecture Release 1.0*. Canada HL7. November 6, 2000.

[2] S. Bakken, K.E. Campbell, J.J. Cimino, S.M. Huff, and W.E. Hammond. "Toward Vocabulary Domain Specifications for Health Level 7-Coded Data Elements", *J Am Med Inform Assoc.* 2000, 7(4): 333-42.

[3] J.H. van Bemmel and M.A. Musen. *Handbook of Medical Informatics*. Springer-Verlag, 1997.

[4] B. Blobel, P. Pharow, K. Engel, V. Spiegel, and R. Krohn. "Communication Security in Open Health Care Networks", *Stud Health Technol Inform.* 1999, 68: 291-6. [Available: http://hi-europe.co.uk/files/1998_9/bridges/comm_sec_open_hcnetworks.htm]

[5] Robin Cover. "Health Level Seven XML Patient Record Architecture". *The XML Cover Pages*. [Available: http://xml.coverpages.org/hl7PRA.html]

[6] Robert H. Dolin. "Clinical Document Architecture", *HL7 International Affiliates Joint Meeting*. August 2000, Dresden, Germany. [Available: http://www.hl7.de/downloads/iajm2000/dolin.pdf]

[7] R. Dolin. "Clinical Document Architecture", *e-Health 2001: The Future of Health Care Proceedings*. 26-29 May 2001, Toronto, Ontario. Canada HL7 Meeting. [Available: http://www.cihi.ca/hl7/may2001/TuteHealth.052001.pdf]

[8] Robert H. Dolin, Liora Alschuler, Sandy Boyer, Calvin Beebe. "An Update on HL7's XML-based Document Representation Standards", *Proc. of the AMIA 2000 Annual Symposium*, November 4-8, 2000, Los Angelos, California. [Available: http://www.amia.org/pubs/symposia/D200113.pdf]

[9] A. Goh. "Java-Based Framework for the Secure Distribution of Electronic Medical Records", *Stud Health Technol Inform.* 1999, 68:582-7.

[10] HL7 Mission Statement. [Available: http://www.hl7.org/about/hl7mission.htm]

[11] Kona Proposal Committee. "The Kona Proposal for Electronic Health Care Records", July 7, 1997. [Available: http://www.hytime.org/ihc97/papers/harding/kona/kona.html]

[12]H. Liu, and C. Friedman. "A Method for Vocabulary Development and Visualization Based on Medical Language Processing and XML", *Proc. of the AMIA 2000 Annual Symposium*, November 4-8, 2000, Los Angelos, California, p. 502-6

[13] G. Paterson, X. Wang, M. Shepherd, C.Watters, and D. Zitner. "Electronic Exchange of Structured Interim Discharge Summaries Using the XML-based Clinical Document Architecture", *e-Health 2001: The Future of Health Care Proceedings*. 26-29 May 2001, Toronto, Ontario. COACH Conference. CD-ROM. [Available: http://www.medicine.dal.ca/dmedinfo/ehealth2001f.ppt]

[14] G. Schadow. "HL7 Version 3 Data Types Ballot". Draft 1. Revision 1.2 [Available: http://aurora.rg.iupui.edu/v3dt/V3DT-ballot-1-print.pdf]

[15] M. Shepherd, D. Zitner, and C. Watters**.** **"**Medical Portals: Web-Based Access to Medical Information". *Proceedings of the 34th Hawaii International Conference on System Sciences*. 4-7 January 2000. IEEE Conference. CD-ROM.

[16] G. Shobowale. "SGML, XML and the Document-Centered Approach to Electronic Medical Records", *Bulletin of the American Society for Information Science*. October/November 1998:7-10.

[17] R. Sokolowski. "XML and its Impact on Content and Structure in Electronic Health Care Documents", *AMIA 1999 Proceedings*. CD-ROM. [Available: http://www.amia.org/pubs/symposia/D005438.PDF]

[18] R. Soper. "A Study of Alternative Storage Strategies for Medical XML Documents", Work Term Report, Faculty of Computer Science, Dalhousie University 2001.

[19] P. S. Woolman, J. Oates, J. Frame, and J. Hardy. "XML for Immediate Discharge Letters in Scotland", *Stud Health Technol Inform.* 2000, 77: 1040-4.

[20] P. S. Woolman, J. Oates, J. Frame, and J. Hardy. "Use of Extensible Markup Language (XML) for Transfer of Immediate Discharge Letter information between Hospital and General Practitioner in Scotland", *Healthcare Computing* 2000. [Available: http://scotland-xml.uk.eu.org/documents/oldpapers/xml-paper-v1.htm]