An XML Portable Chart Format

Henry C. Chueh M.D. M.S., Wayne F. Raila, David A. Berkowicz M.D.,
and G. Octo Barnett M.D.
Laboratory of Computer Science,
Massachusetts General Hospital, Boston, MA.

ABSTRACT

The clinical chart remains the fundamental record of outpatient clinical care. As this information migrates to electronic form, there is an opportunity to create standard formats for transmitting these charts. This paper describes work toward a Portable Chart Format (PCF) that can represent the relevant aspects of an outpatient chart. The main goal of the format is to provide a packaging medium for outpatient clinical charts in a transfer of care scenario. A secondary goal is to support the aggregation of comparable clinical data for outcomes analysis. The syntax used for PCF is Extended Markup Language (XML), a W3C standard. The structure of the PCF is based on a clinically relevant view of the data. The data definitions and nomenclature used are based primarily on existing clinical standards.

INTRODUCTION AND BACKGROUND

Health care consumers are increasingly mobile. Transfer of clinical care is a common occurrence for many patients. Permanent transfers of care occur when patients change primary care physicians. Temporary transfers occur through the referral and consultation processes. Typically, if any clinical records are transferred, paper charts are copied and then mailed or faxed. Many times the copied chart does not find its way to the clinician in a timely manner. In part because of these issues, an increasing number of health care sites are adopting or considering the use of electronic records. Despite this, transfer of this clinical data between even these sites will occur usually through paper printouts, since there is no single accepted standard for the electronic representation of outpatient clinical charts. Some solutions to construct virtual records using existing clinical standards have been successful[1,2].

Efforts to standardize health care information are not new. One of the most successful and longstanding initiatives is Health Level 7 (HL7)[3]. Its primary goal is the electronic exchange of information between different systems within institutions. Transactions involving patient admission and discharge events, clinical encounters, billing, orders, and results of laboratory tests are represented. Other standards initiatives in health care informatics include a variety of efforts at the American Society of Testing and Materials (ASTM)[4]. Both HL7 and CorbaMed[5], a task force of the Object Management Group, have initiatives for creating standardized object-oriented models for data interchange.

These and other standards initiatives are promising approaches for electronic sharing of health care information. Most have comprehensive, broad-based aims involving transmission of more than the basic patient clinical data as represented in the outpatient clinical chart. Some efforts have a goal of defining a comprehensive information model for health care. On the other hand, there is a current need for a simple format to package a clinical chart for electronic transmission. This need has led to the Portable Chart Format (PCF) initiative described here. The intent of the PCF is to provide an extensible format around which outpatient electronic medical record (EMR) applications can be designed. Use of a portable format can promote both site-independent records as well as aggregation of data for outcomes [1,2,6].

Particular features of PCF that were considered to be valuable included 1) a simple, readable format, 2) the ability to validate the format in an automated fashion, 3) site and database independence, and 4) consideration of data aggregation for outcomes. PCF is not meant to provide a format for all things transmissible about the health care domain. Nor is it expected that the technologies used in PCF would be appropriate for all health care systems.

DESIGN AND IMPLEMENTATION

Design

The approach we have taken to defining a Portable Chart Format can be divided into four specific areas:

1) Definition of scope and assumptions
2) Consistency with existing standards
3) Technology selection
4) Process of defining PCF

Scope. The scope of this initiative is to define a solution for the electronic portability of a typical outpatient clinical chart. An assumption is made that

1091—8280/98/$5.00

this information is available electronically from databases. For data not stored as ASCII text, these data would need to be available as digital images or scanned documents.

Two scenarios that help to define the boundaries of the scope are: 1) a transfer of patient care between primary care physicians, and 2) aggregation of clinical data across patients for analysis. The first scenario requires a data structure that can hold all relevant contents of an outpatient office chart. This implies data structures that can handle diverse entries from visit notes to an active medication list. The second scenario implies a format that can express clinical data with varying degrees of granularity.

As important are the issues that the PCF does not address. While the PCF does make some assumptions about the types or classes of data that form a typical clinical chart, it does not attempt to define a comprehensive data model. In this respect the PCF is as much a packaging metaphor as an information representation. The only requirement is that elements of a specific type are placed in a consistent location in the package.

The relationship between the data elements that would constitute a more complete data model is neither expressed nor excluded. A more comprehensive data model is likely to exist for any particular electronic medical record application, but PCF does not reflect this directly. The reason for this approach is twofold. First, an expectation that a single comprehensive data model is appropriate for all EMR databases is probably not realistic. Second, it is likely that site-specific data models will evolve over time to support site-specific clinical measures and workflow, and the format for transmitting clinical data should not limit this by imposing a fixed data model.

Standards. Leveraging existing standards efforts while creating a simple, open format is a goal. Existing data definitions and nomenclatures are being used whenever possible. A new data definition is defined only under the following criteria: 1) the data is critical to the scope of the format, and 2) it has not been defined in accepted standards. For example, HL7 has many elements of a problem defined. While PCF repackages these through XML data structures, many of the data fields have direct correspondence with HL7. Conversely, not all PCF data elements for clinical data are included in HL7 (Table 1).

XML	HL7 2.3	
<PROBLEM> subelement	segment	name
CONCEPT.ID	PRB-3.1	Problem ID <identifier>
CONCEPT.TEXT	PRB-3.2	Problem ID <text>
CONCEPT.SCHEME	PRB-3.3	Problem ID <scheme>
INSTANCE	PRB-4.1	Problem Instance <identifier>
	PRB-5	Episode of Care ID
CATEGORY.ITEMORDER	PRB-6	Master problem list #
	...	
RESOLUTIONDATE	PRB-9	Actual problem resolution date/time
CATEGORY	PRB-10	Problem class.
	PRB-11	Problem mgmt disc.
MODIFIER	PRB-12	Problem persistence
MODIFIER	PRB-13	Problem confirmation
MODIFIER	PRB-14	Problem life cycle
	PRB-15	Problem life cycle status date/time
ONSETDATE	PRB-16	Problem date of onset
	PRB-17	Problem onset text
CATEGORY	PRB-18	Problem ranking
	...	
SENSITIVITY	PRB-25	Security/sensitivity
XML tags not in HL7 2.3		
DESCRIPTION		Brief narrative synopsis of problem

Table 1. Partial mapping of HL7 data definitions to PCF XML for <PROBLEM> element

Technology. Since the intent of PCF is to be site and database independent, no specific software or hardware standards are specified. Instead, technology is limited to the selection of a message syntax. Extended Markup Language (XML) was chosen because of its flexibility and design for use with World Wide Web (Web) technologies. The rapid adoption of Web-based technologies is evident[7,8], and use of XML takes advantage of this infrastructure. Previous investigation in the health care domain with the parent of XML, Structured Generalized Markup Language (SGML), has been promising[9]. XML has been adopted as a formal World Wide Web Consortium (W3C) standard[10].

Process. Ultimately an open consensus process for defining PCF is necessary. In this initial work, we rely on experience with real-world EMR systems and the expectations of clinicians within the clinical environment. Creating applications that actually use a preliminary PCF can help to validate and guide the process further. Ideally these applications should view and manipulate data interchanged with PCF during actual clinical encounters.

Implementation

An implementation against three disparate databases was used to exercise the preliminary versions of the PCF. They include two active electronic medical records at Massachusetts General Hospital – CoSTAR[11] ('M database), and the EMR[12] (Oracle database) -- and the Boston Health Care for the Homeless Program (BHCHP) EMR[13], also Oracle-based. These three databases all support outpatient records but have different data fields and data models. Services to generate PCF charts from each of these databases were created. An architecture for retrieving XML-based charts was established, and a Web-based electronic record was created to visualize and manage those charts. The latter effort was performed in the context of designing and developing a next generation EMR for the Boston Health Care for the Homeless Program.

RESULTS

An initial skeleton of PCF has been defined which encompasses a prototype clinical chart. Figure 1 shows an example of a partial skeleton PCF with some of the major section elements.

```
<PCF>
  <HEADER>
    <SENDINGFACILITY></SENDINGFACILITY>
    <RECEIVINGFACILITY></RECEIVINGFACILITY>
  </HEADER>
  <CHART>
    <PATIENT>
      <ID TYPE= FACILITY=></ID>
      <PERSONNAME>
       <LASTNAME></LASTNAME>
       <FIRSTNAME></FIRSTNAME>
      </PERSONNAME>
      <SEX ID= SCHEME=></SEX>
      <DOB></DOB>
    </PATIENT>
    <PROBLEMLIST></PROBLEMLIST>
    <ALLERGYLIST></ALLERGYLIST>
    <MEDLIST></MEDLIST>
    <SOCIALHX></SOCIALHX>
    <NOTES></NOTES>
    <LABS></LABS>
  </CHART>
  <CHART>
  </CHART>
</PCF>
```

Figure 1. PCF skeleton (simplified)

A single PCF document can hold multiple charts. A PCF document has a single sender and recipient, even though each chart contained within it can be from different sites. Using PCF is similar to sending an electronic filing cabinet that contains multiple charts. Most elements in a PCF document are optional. The few mandatory elements pertain to the identification of each patient chart as well as the sender.

PCF explicitly recognizes two distinct classes of data that are relevant in an outpatient clinical chart. One class is chronological document-based data such as visit notes. The other class is summary or "snapshot" data such as a problem list or medication list. Some would consider that summary data is derived from the chronological data. For example, problems on a problem list may be identified in visit notes. However, this is not always the case, as many EMR systems allow problem lists to be maintained separately from visit notes[14]. Even if possible, retrieving a chart that includes all visit notes just to derive a problem list is not efficient. Figure 2 shows detail for the PROBLEMLIST element (Figure 5 shows this XML rendered visually as HTML).

```
<PROBLEMLIST>
  <PROBLEM>
    <INSTANCE>49956</INSTANCE>
    <TERM>
      <NAME>Contact dermatitis</NAME>
      <CONCEPTID>1211</CONCEPTID>
      <SCHEME>HCH</SCHEME>
    </TERM>
    <MODIFIER TYPE="STATUS">ACTIVE</MODIFIER>
    <CATEGORY TYPE="" ITEMORDER=""></CATEGORY>
    <DESCRIPTION></DESCRIPTION>
    <ONSETDATE>
      <YEAR>1998</YEAR>
      <MONTH>06</MONTH>
      <DAY>16</DAY>
    </ONSETDATE>
    <RESOLUTIONDATE></RESOLUTIONDATE>
    <SENSITIVITY></SENSITIVITY>
    <OBSERVATION>
      <INSTANCE>248907</INSTANCE>
      <MADEBY>
        <ID TYPE="" FACILITY=""></ID>
        <PERSONNAME> </PERSONNAME>
      </MADEBY>
      <TEXT><![CDATA[History of contact dermatitis. Reports allergie's to
detergent and certain soaps (IVory, Zest, Dial). No rash...]]></TEXT>
      <SERVICEDATE> </SERVICEDATE>
      <SIGNATURE></SIGNATURE>
    </OBSERVATION>
  </PROBLEM>
  <PROBLEM>
    <INSTANCE>56093</INSTANCE>
    <TERM>
      <NAME>PPD negative</NAME>
      <CONCEPTID>12660</CONCEPTID>
      <SCHEME>HCH</SCHEME>
    </TERM>
    <MODIFIER TYPE="STATUS">ACTIVE</MODIFIER>
    <OBSERVATION>
      <INSTANCE>85759</INSTANCE>
      <TEXT><![CDATA[1996. Will repeat today. ]]></TEXT>
    </OBSERVATION>
    ...
  </PROBLEM>
  ...
</PROBLEMLIST>
```

Figure 2. Simplified partial PCF Problem List

XML has a Document Type Definition (DTD) which specifies the rules or grammar of how a particular XML document type must be constructed. Figure 3 shows a portion of the PCF DTD that defines a problem list. A preliminary framework for PCF requests has also been defined. There are a number of existing generalized query languages. Structured Query Language (SQL) is one example. SQL presumes an underlying relational data schema. Other languages are independent of a logical schema, but can be complex because of the need to handle generic queries. Instead, we adopted a model of query-by-example (QBE) to make PCF requests.

```
<!ELEMENT PROBLEMLIST    ( PROBLEM*, SSPROBLEMLIST?)                    >

<!ELEMENT  PROBLEM       (INSTANCE,    TERM,    MODIFIER*,
                          CATEGORY*,          DESCRIPTION?,
                          ONSETDATE?,  RESOLUTIONDATE?,
                          SENSITIVITY?,       OBSERVATION*,
                          SSPROBLEM?)                              >
<!ELEMENT  TERM          (#PCDATA)>
<!ATTLIST  TERM
           ID            CDATA                    #IMPLIED
           SCHEME        CDATA                    #IMPLIED       >
<!ELEMENT  CATEGORY      (#PCDATA)                                  >
<!ATTLIST  CATEGORY
           TYPE          CDATA                    #IMPLIED
           ITEMORDER     CDATA                    #IMPLIED       >
<!ELEMENT  MODIFIER      (#PCDATA | (CODE, SCHEME?))*               >
<!ATTLIST  MODIFIER
           TYPE          (%ModifierType;)      #REQUIRED
           ID            CDATA                    #IMPLIED
           SCHEME        CDATA                    #IMPLIED       >
<!ELEMENT  DESCRIPTION   (#PCDATA)                                  >
<!ELEMENT  ONSETDATE     %Time;                                     >
```

Figure 3. A portion of the PCF DTD for the PROBLEMLIST element

A request is made up of PCF itself with the elements desired included with attributes to qualify aspects of the query. In order to simplify the query for a complete chart request, a <FULLCHART> tag is used to indicate all available PCF sections. Figure 4 shows a simplified request for two separate charts. The first chart is requested by medical record number (MRN), confirmed with name and sex, and includes problem and medication lists, and visit notes since 1995. The other chart is retrieved by Social Security Number (SSN) and includes all available sections.

```
<PCF>
  <HEADER></HEADER>
  <CHART>
    <PATIENT>
      <ID TYPE="MRN">123-45-67</ID>
      <PERSONNAME>
       <LASTNAME>Doe</LASTNAME>
       <SEX ID="M" SCHEME="HL7">M</SEX>
      </PERSONNAME>
    </PATIENT>
    <PROBLEMLIST></PROBLEMLIST>
    <MEDLIST></MEDLIST>
    <NOTES STARTDATE="01/01/1995"></NOTES>
  </CHART>
  <CHART>
    <PATIENT>
      <ID TYPE="SSN">123-45-6789</ID>
    </PATIENT>
    <FULLCHART STARTDATE="01/01/1994"></FULLCHART>
  </CHART>
</PCF>
```

Figure 4. Sample PCF query (simplified).

The advantage of QBE is that the granularity of the query can be established easily through the exclusion or inclusion of specific PCF XML tags.

A Web-based EMR application was constructed that can retrieve and parse XML documents within a Web browser. This application was then directed to the three separate PCF generating services for the EMR databases discussed above. Each service is composed of an Active Server Page (ASP) script running on a Microsoft IIS Web server which retrieves data from the database and formats it into the PCF XML format. These services are exposed as individual Web server addresses (URL) which accept a request and return a PCF XML document. When the XML arrives at the client, it is parsed using the freely available Microsoft Java-based XML parser. The resulting XML tree is then rendered into HTML and presented (Figure 5).

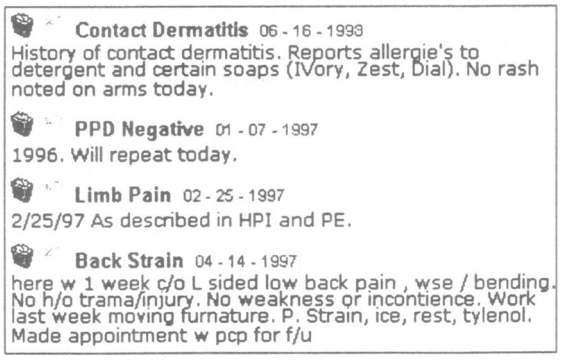

Figure 5. HTML rendering of a PCF problem list.

DISCUSSION

An argument can be made that there is no need for another standard for the transmission of clinical data. In fact, new standards such as the Web, HTML and XML have created opportunities that are not yet leveraged fully by existing standards. Building on existing standards and incorporating them into a new, international, cross-industry standard (XML) opens the standard to a broader audience. It may also increase the opportunities for successful dissemination and implementation of the standard. The use of XML as a syntax for a portable clinical chart format has let us create a model for the transfer of clinical charts. As an adopted W3C standard, an increasing amount of commercial software to manage XML data is expected. The PCF effort therefore has been able to leverage existing standards in both software as well as content.

The scalability of PCF is a major issue. First, the ability to transfer varying amounts of clinical data has importance. This is provided through the optional nature of PCF tags, the QBE request model, and time ranges that can be specified with different parts of each PCF request. The other issue is that of data granularity. For example, one site may have many attributes associated with each problem on the problem list, while another site may simply have the clinical diagnosis. Or, one site may have discrete medication fields such as dose and frequency, while another site has a prescription field that is composed of all this data together. Supporting these variations

of scale is critical to the success of any packaging format such as PCF. The fact that XML allows data elements to be specified formally as optional solves much of this problem. The ability to add site-specific tags provides additional flexibility.

The relative lack of semantic data relationships and logic within PCF -- the inability, for instance, to directly relate a problem to a note in which it was mentioned -- has certain drawbacks. However, many of these relationships can be derived at the application level. Also, site-specific tags can be used to store semantic information that is only relevant for certain databases. Keeping the data model issues minimal has allowed a more rapid definition of a straightforward format. When comprehensive data models of the health care domain emerge that become a standard of practice, PCF should be revised to accommodate these changes.

The current working version of PCF does not incorporate any proprietary security algorithms. For data security it relies on secure transmission of XML documents using already available technologies. Use of PCF in any system will still require the appropriate implementation of confidentiality and access policies. If to be extended, this initial work with PCF should be discussed through an open consensus standards process. For example, a newly formed ASTM subcommittee E31.11 on Electronic Health Record Portability may consider such formats. Additional implementations with diverse databases will be helpful to expose potential site-specific problems with PCF. Studies of actual chart migration and data aggregation using the format will also be needed.

CONCLUSION

The creation of an XML-based format for a portable clinical chart is one viable approach for site-independent electronic clinical records. Prototype implementations of the format against different production databases have been promising. Implementing PCF has allowed us to plug different data sources into an existing application framework and have instant access to the data. A clinically meaningful chart can be transmitted to any PCF compliant application without any prior negotiation between the sender and the recipient. As XML gains momentum on the Web and in the health care domain, the opportunity for cross-fertilization with other XML-based initiatives is expected.

Acknowledgements

This work was supported in part by NLM research Grant LM05854 and a research grant from Hewlett Packard Corporation.

References

1. Kohane IS, van Wingerde FJ, Fackler JC, Cimino C, Kilbridge P, Murphy S, et al. Sharing electronic medical records across multiple heterogeneous and competing institutions. Proc AMIA Annu Fall Symp. 1996; : 608-612.
2. Halamka JD, Safran C. Virtual consolidation of Boston's Beth Israel and New England Deaconess Hospitals via the World Wide Web. Proc AMIA Annu Fall Symp. 1997; : 349-353.
3. Health-Level 7 Standards. Available from: URL: http://www.mcis.duke.edu/ standards/ HL7/ hl7frontpage.htm.
4. ASTM Committee E-31 on Healthcare Informatics. Available from: URL: http://www.astm.org/commit/e-31.htm
5. CorbaMed. Available from: URL: http://www.omg.org/corbamed/corbamed.htm
6. Dolin RH. Outcome analysis: considerations for an electronic health record. MD Comput. 1997 Jan; 14(1): 50-56.
7. Lowe HJ, Lomax EC, Polonkey SE. The World Wide Web: a review of an emerging internet-based technology for the distribution of biomedical information. J Am Med Inform Assoc. 1996 Jan; 3(1): 1-14.
8. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. J Am Med Inform Assoc 1995 Sep;2(5):273-284.
9. Dolin RH, Alschuler L, Bray T, Mattison JE. SGML as a message interchange format in healthcare. Proc AMIA Annu Fall Symp. 1997: 635-639.
10. World Wide Web Consortium. The World Wide Web Consortium issues XML 1.0 as a W3C recommendation: XML 1.0 fact sheet 1998. Available from: URL: http://www.w3.org/Press/1998/XML10-REC-fact.html
11. Barnett GO. The application of computer-based medical-record systems in ambulatory practice. N Engl J Med. 1984 Jun 21; 310(25): 1643-1650.
12. Chueh HC, Raila WF, Pappas JJ, Ford M, Zatsman P, Tu J, et al. A component-based, distributed object services architecture for a clinical workstation. Proc AMIA Annu Fall Symp. 1996; : 638-642.
13. Chueh HC, Barnett GO. Client-server, distributed database strategies in a health-care record system for a homeless population. J Am Med Inform Assoc. 1994 Mar; 1(2): 186-198.
14. Kuperman G, Bates DW. Standardized coding of the medical problem list. J Am Med Inform Assoc. 1994 Sep; 1(5): 414-415.