

HL7 Document Patient Record Architecture: An XML Document Architecture Based on a Shared Information Model

Robert H. Dolin, MD¹; Liora Alschuler²; Fred Behlen, PhD³; Paul V. Biron, MLIS¹;
Sandy Boyer, RPh; Dan Essin, MD⁴; Lloyd Harding⁵; Tom Lincoln, MD^{6,7}; John E.
Mattison, MD¹; Wes Rishel⁸; Rachael Sokolowski⁹; John Spinosa, MD, PhD¹⁰; Jason P.
Williams, MS¹¹

¹Kaiser Permanente; ²The Word Electric; ³U. of Chicago; ⁴USC; ⁵Information Assembly
Automation Inc.; ⁶Rand Corp.; ⁷U. of Illinois at Chicago; ⁸Wes Rishel Consulting;
⁹Magnolia Technologies; ¹⁰Scripps Health, PMG; ¹¹Oceania Inc.

The HL7 SGML/XML Special Interest Group is developing the HL7 Document Patient Record Architecture. This draft proposal strives to create a common data architecture for the interoperability of healthcare documents. Key components are that it is under the umbrella of HL7 standards, it is specified in Extensible Markup Language, the semantics are drawn from the HL7 Reference Information Model, and the document specifications form an architecture that, in aggregate, define the semantics and structural constraints necessary for the exchange of clinical documents. The proposal is a work in progress and has not yet been submitted to HL7's formal balloting process.

INTRODUCTION

In 1996, the HL7 SGML initiative evolved as a special interest group (SIG) of HL7¹. The HL7 SGML/XML SIG is interested in coordinating the development of a comprehensive document architecture for healthcare; educating the healthcare community in the capabilities and utility of XML-based information; developing, coordinating, and maintaining a framework for the interoperability of healthcare documents; coordinating and cooperating with other XML initiatives; and investigating the use of XML as a messaging syntax². The proposal being developed by the HL7 SGML/XML SIG to address the interoperability of healthcare documents is known as the HL7 Document Patient Record Architecture (PRA)³.

The need for a patient record architecture stems from the desire to unlock the considerable clinical content currently stored in free text clinical notes, and to enable pooling of content from documents created on systems of widely varying characteristics. Given the variability in clinical notes, including structure, underlying information models, degree of semantic

encoding, use of standard healthcare terminologies, platform- and vendor-specific features, it is currently difficult to store and/or exchange documents with retention of computer-processable semantics over both time and distance. The idea that a single healthcare document XML DTD could be developed and broadly implemented seems unrealistic. The HL7 SGML/XML SIG proposes that what is needed is a common data architecture that can accommodate a diverse set of records and requirements. The PRA strives to be such an architecture.

The PRA is an XML-based architecture. XML (Extensible Markup Language) (www.w3.org/TR/1998/REC-xml-19980210.html) reduces a document to a word in a known context-free grammar through a process of markup. The formal markup specification for a collection of documents is called a Document Type Definition (DTD). Documents are then written to conform to a particular DTD, enabling them to be automatically parsed and validated against that DTD. XML is a proper subset of SGML (Standard Generalized Markup Language, ISO 8879:1986), meaning that all valid XML documents are SGML documents. For more on SGML/XML, see the references^{4,6}.

The DTDs put forth by the HL7 SGML/XML SIG will be harmonized with the evolving HL7 Reference Information Model (RIM)⁷. This information model will serve as the central schema defining the semantics for all HL7 messages and documents. Fields in all HL7 messages will map to the RIM. Likewise, components in PRA DTDs will map to the RIM.

The PRA "architecture" is envisioned as a hierarchically organized set of document schemas that, in aggregate, define the semantics and structural constraints necessary for the

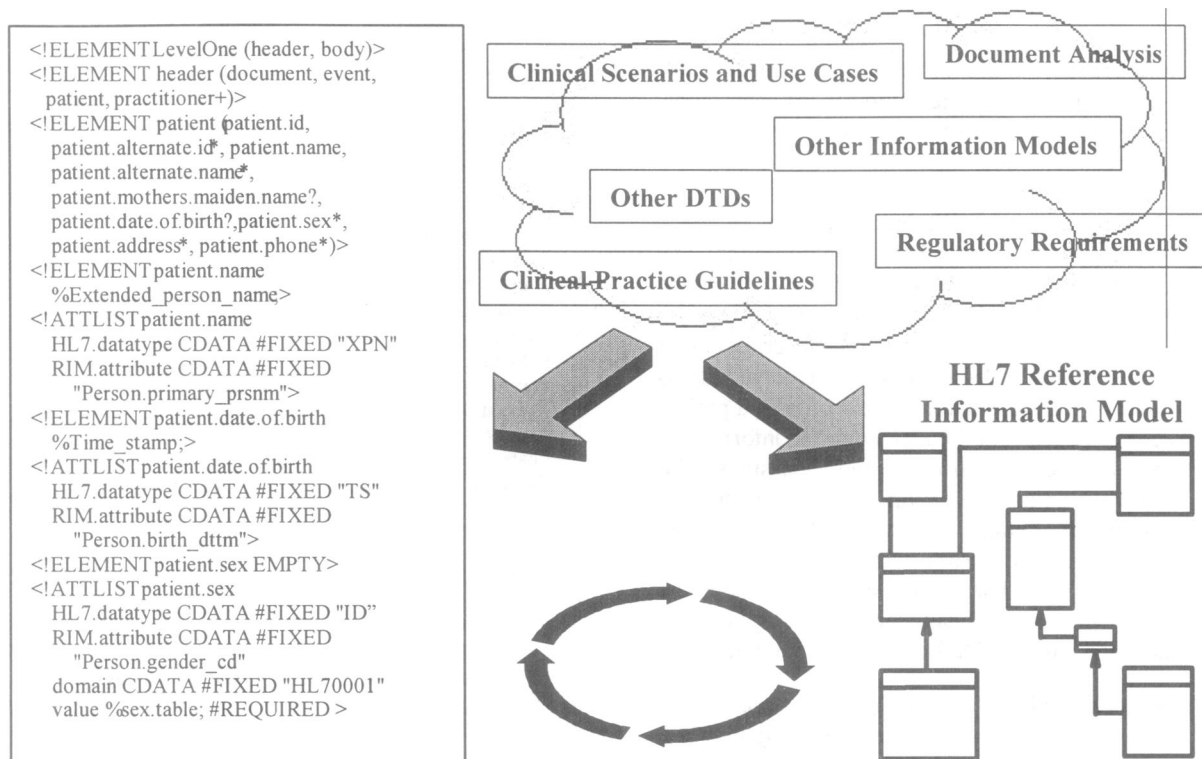


Figure 1. PRA DTDs are harmonized with the HL7 RIM

exchange of patient record documents.

The PRA is a work-in-progress. While many of the fundamental principles have been discussed and documented, and drafts of the foundational DTDs have been created, the material presented in this report should be considered the opinions of the authors, and not a consensus within HL7 or any group within HL7. The material in this paper will ultimately be subjected to HL7's formal balloting process.

PRA OVERVIEW

Scope

The scope of PRA is the exchange of clinical documents. The definition of "clinical document" is evolving, but includes those documents generated by healthcare practitioners documenting the care provided for individual patients. The architecture specifies a document markup format which can be transferred on-line or stored in files on off-line media. An HL7 Document is a defined and persistent information object that can exist outside of a messaging context.

Principles

One of the foundational principles of PRA is that it be harmonized with the RIM (Figure 1). The requirements for the semantic content of PRA are derived from numerous sources. The harmonization process is an iterative cycle of

feeding identified requirements into the RIM, ensuring that the RIM reflects these requirements, and mapping PRA components to the RIM. In Figure 1, the fixed XML attribute "RIM.attribute" expresses the mapping from an XML element to a corresponding RIM attribute. The fixed XML attribute "HL7.datatype" expresses the HL7 data type of the XML element. (HL7 data types are defined as entities within the DTD and referenced by other XML elements.)

The PRA is envisioned as a hierarchically organized set of document schemas (or DTDs) that, in aggregate, define the semantics and structural constraints necessary for the exchange of clinical documents. The proposal defines a multi-level document architecture where each level is derived from a more basic level, with level one being the most basic. Higher levels enable the computer-processable expression of richer shared semantics. We anticipate domain-specific specializations branching out from each level (such as a Cardiology DTD branching out from the generic PRA level two DTD).

The PRA levels can be thought of as levels of conformance. Regulatory agencies can require a particular PRA level for reporting. Institutions with different abilities to generate and process XML-encoded documents, can claim conformance with a particular PRA level. This

should enable more institutions to utilize PRA while maximizing the amount of computer-processable and shared semantics embedded in an exchanged document.

Level one (a.k.a. "Coded Header") specifies encoding of the clinical document header. The document body can be non-XML data, or can be encoded using largely structure-based tags. An example (not valid per the level one DTD in that some required elements are not included) is shown in Figure 2. Every level contains a patient record header. The header contains information that uniquely identifies and classifies the document; plus attestation, event, patient, and practitioner details. Every level can contain coded entries from standard healthcare terminologies. These are expressed with the XML element "healthcare.code" which is patterned after the HL7 coded element data type.

Level two (a.k.a. "Coded Context") contains the same header as in level one, and structures the document body into coded sections. The vision for level two is that it will provide the semantics to encode the context under which clinical events occur.

Level three (a.k.a. "Coded Content") is envisioned to encode a document sufficient to meet the processing needs of a fully electronic health record. Figure 3 is an example of how such detailed semantics might be represented in an XML document. The example is taken from an earlier study that examined various ways of representing RIM concepts within an XML document⁸, and does not reflect a draft PRA DTD. The point of the example is not to suggest that this is the correct model, but to show that it is possible to reflect the detailed semantics of an information model such as the RIM within an XML document. Within the "impressions" section, the statement "Nodule in the RLL, suggestive of malignancy" is semantically encoded using RIM (version 0.84) objects "Clinical_observation" and "Observation_relationship".

DISCUSSION

A shared information model is central to PRA. HL7 messages will also be based on the RIM, enabling the interplay of PRA documents and HL7 messages. Such interplay may include deriving an order message from the Plan section

```

<LevelOne>
<header>
<document>
  <document.id>
    <id.value>CXR001</id.value>
  </document.id>
  <doc.creation.date>19991101</doc.creation.date>
  <document.type>
    <identifier>P5-00010</identifier>
    <text>DiagnosticRadiologic Examination</text>
    <name.of.code.system>SNM3</name.of.code.system>
  </document.type>
</document>
<event>
  <event.date>19991101</event.date>
</event>
<patient>
  <patient.id>
    <id.value>1234789</id.value>
  </patient.id>
  <patient.name>
    <family.name>Levin</family.name>
    <given.name>Henry</given.name>
    <suffix>the 7th</suffix>
  </patient.name>
  <patient.date.of.birth>19230113</patient.date.of.birth>
  <patient.sex.value="male"/>
</patient>
<practitioner>
  <practitioner.id>
    <id.value>24680</id.value>
  </practitioner.id>
  <family.name>Fall</family.name>
  <given.name>Amy</given.name>
  <mi>A</mi>
  <prefix>Dr.</prefix>
</practitioner>
</header>
<body>
<section>
  <section.title>Procedure</section.title>
  <paragraph>
    <healthcare.code.identifier="P5-20100"
      name.of.coding.system="SNM3"
      local.coding.system="N">ChestX-Ray
    </healthcare.code>
  </paragraph>
</section>
<section>
  <section.title>Findings</section.title>
  <paragraph>RLL nodule</paragraph>
</section>
<section>
  <section.title>Impressions</section.title>
  <paragraph>Nodule in the RLL. suggestive of
    malignancy.</paragraph>
</section>
<section>
  <section.title>Recommendations</section.title>
  <paragraph>I notified the ordering physician of this
    finding.</paragraph>
</section>
</body>
</LevelOne>

```

Figure 2. Sample PRA level one document.

of a document, importing a lab report into the Results section of a document, and the ability to send documents as messages and communicate with systems designed for message storage, indexing, and retrieval.

A distinction can be made between an exchange DTD and an authoring DTD. The PRA proposes an architecture for the exchange of clinical documents. An HL7 PRA document may be originally authored data (meaning that an institution adopts a PRA DTD as their internal DTD), or may be a transformation from original data (meaning that an institution may have their own internal DTD, and map into the PRA for purposes of exchange). In the latter case, the exchanged document is not necessarily the same as the originally attested document.

Given this distinction, and because many institutions have or are actively developing their own internal document representations, there is a need to map or transform from an institution's authoring DTD into a PRA exchange DTD. The architectural relationship of DTDs comprising the PRA are currently specified with the Architectural Forms Definition Requirements (AFDR, ISO/IEC 10744). AFDRs also specify the processing semantics of "architectural engines" which transform document instances from one DTD to another. Other SGML/XML transformation languages include the Document Style Semantics and Specification Language (DSSSL, ISO/IEC 10179:1996) (www.jclark.com/dsssl/) and the draft Extensible Stylesheet Language (XSL) (www.w3.org/Style/XSL/). General programming languages also enable the transformation from one DTD to another.

It is important to note that each transformation language has potential syntactic limitations in the type of mappings that can be formally expressed. Additionally, there are semantic mappings that cannot be resolved regardless of the transformation language employed. While the discussion here focuses on mapping from one DTD to another, the issues are no different than those that arise in the more general sense of mapping from one information model to another^{9,10}. Figure 4 represents an institution's local document. The institution wants to map this document into a PRA level three document (such as the one in Figure 3). Elements in the source DTD may be in different order than in a PRA DTD (e.g. <dob> <name> vs. <patient.name>

```

<LevelThree>
<header>
<!-- same as LevelOne header-->
</header>
<body>
<procedure>
<paragraph>
<healthcare.code identifier="P5-20100"
name.of.coding.system="SNM3"
local.coding.system="N">ChestX-Ray
</healthcare.code>
</paragraph>
</procedure>
<findings>
<paragraph>RLL nodule</paragraph>
</findings>
<impressions>
<Clinical_observation>
<observation_value_txt code="M-03010"
source="SNM3"/>Nodule
<Observation_relationship>
<relationship_type_cd code="GC006"
source="SNM3"/>in the
<Clinical_observation>
<observation_value_txt code="T-28400"
source="SNM3"/>RLL,
</Clinical_observation>
</Observation_relationship>
<Observation_relationship>
<relationship_type_cd code="G-C022"
source="SNM3"/>suggestive of
<Clinical_observation>
<observation_value_txt code="M-80001"
source="SNM3"/>malignancy.
</Clinical_observation>
</Observation_relationship>
</Clinical_observation>
</impressions>
<recommendations>
<paragraph>I notified the ordering physician of this
finding.</paragraph>
</recommendations>
</body>
</LevelThree>

```

Figure 3. Representation of detailed RIM semantics within an XML document.

<patient.date.of.birth>). The source DTD may be missing a required element (e.g. PRA DTD requires <document.id>). The source and PRA DTDs may have different enumerated list values (e.g. <sex>M</sex> vs. <patient.sex value="male"/>). Elements may have different data types (e.g. <dob>January 13, 1923</dob> vs. <patient.date.of.birth>19230113</patient.date.of.birth>). The source DTD may have coarser granularity than the PRA DTD (e.g. <name>Henry Levin, the 7th</name> vs. <patient.name><family.name>Levin</family.name><given.name>Henry</given.name><suffix>the 7th</suffix></patient.name>).

Perhaps the most challenging mapping problem occurs when the source's information model of

the real world differs from the information model underlying PRA. Figure 3 embeds concept and relationship semantics within the text of a sentence, using a nested representation. In Figure 4 the tags embedded in the practitioner-specified text serve as anchors, and the concepts in that text along with the relationships between those concepts are specified after the text, making references back to the anchors (somewhat similar to that described by Friedman, et al¹¹). A more subtle source of error arises when only a partial semantic mapping is possible, such as when a source document stating "zestril resulted in a severe reaction", is transformed into a document only conveying shared semantics for "zestril".

A conclusion that can be drawn is that mapping to higher levels of the PRA may depend on the extent to which the sender's information model can map to the RIM, or on the extent by which the author DTD can map to a PRA DTD, thus

```

<radiology.report>
<header>
<date.of.creation>November 11, 1999</date.of.creation>
<date.of.study>November 11, 1999</date.of.study>
<mrn>123456789</mrn>
<dob>January 13, 1923</dob>
<name>Henry Levin, the 7th</name>
<sex>M</sex>
<practitioner.id>24680</practitioner.id>
<practitioner.name>Dr. Amy A. Fall</practitioner.name>
</header>
<body>
<section title="Procedure">
<user.text>Chest X-Ray</user.text>
</section>
<section title="Findings">
<user.text>RLL nodule</user.text>
</section>
<section title="Impressions">
<user.text><c v="1"/>RLL <c v="2"/>nodule, <c v="3"/>
suggestive of <c v="4"/> malignancy.</user.text>
<concept c="1" code="T-28400" text="RLL"
source="SNM3"/>
<concept c="2" code="M-03010" text="nodule,NOS"
source="SNM3"/>
<concept c="4" code="M-80001" text="malignancy"
source="SNM3"/>
<relationship c="NULL" c1="2" c2="1" code="G-C006"
text="location" source="SNM3"/>
<relationship c="3" c1="2" c2="4" code="G-C022"
text="suggests" source="SNM3"/>
</section>
<section title="Recommendations">
<user.text>I notified the ordering physician of this
finding.</user.text>
</section>
</body>
</radiology.report>

```

Figure 4. A local XML document to be mapped into PRA.

ultimately to the RIM. Organizations creating source DTDs should consider deriving them from the PRA or mapping constructs directly to the RIM. This holds for groups within HL7, other standards organizations, professional and regulatory groups, providers and vendors. Because the RIM is itself in evolution, the inability to map local information into PRA or the RIM suggests an opportunity for enhancing the shared information model.

References

1. HL7 SGML/XML Special Interest Group website. (www.mcis.duke.edu/standards/HL7/committees/sgml/)
2. Dolin RH, Rishel W, Biron PV, Spinosa J, Mattison JE. SGML and XML as interchange formats for HL7 messages. JAMIA Fall Symp Suppl 1998:720-4.
3. HL7 Document Patient Record Architecture. Draft, Sept, 1998. (www.mcis.duke.edu:80/standards/HL7/committees/sgml/WhitePapers/Prap/)
4. James Clark's SGML web site. (www.jclark.com/)
5. Alschuler L, Dolin RH, Spinosa J. SGML in healthcare information systems. Proc GCA SGML/XML, 1997.
6. Essin DJ. Intelligent processing of loosely structured documents as a strategy for organizing electronic health care records. Meth Inf Med 1993;32:335.
7. HL7 Reference Information Model (www.mcis.duke.edu/standards/HL7/data-model/HL7/modelpage.html)
8. Dolin, RH. HL7 Reference Information Model representation in an XML-based document architecture. Draft. June, 1998. (www.mcis.duke.edu/standards/HL7/committees/sgml/WhitePapers/RIM/xml_rim.rtf)
9. Campbell KE, Oliver DE, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. JAMIA 1998; 5(5): 421-31.
10. CEN/TC 251/PT29. Electronic Healthcare Record Communication. Part 4 - Messages for the exchange of information. Draft, Dec, 1998. (www.clinical-info.co.uk/)
11. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the Extensible Markup Language. JAMIA 1999;6(1):76-87.