

JAMIA

The Practice of Informatics

Review ■

The HL7 Clinical Document Architecture

ROBERT H. DOLIN, MD, LIORA ALSCHULER, CALVIN BEEBE,
PAUL V. BIRON, MLIS, SANDRA LEE BOYER, DANIEL ESSIN, MD,
ELLIOT KIMBER, TOM LINCOLN, MD, JOHN E. MATTISON, MD

Abstract Many people know of Health Level 7 (HL7) as an organization that creates health care messaging standards. Health Level 7 is also developing standards for the representation of clinical documents (such as discharge summaries and progress notes). These document standards make up the HL7 Clinical Document Architecture (CDA). The HL7 CDA Framework, release 1.0, became an ANSI-approved HL7 standard in November 2000. This article presents the approach and objectives of the CDA, along with a technical overview of the standard.

The CDA is a document markup standard that specifies the structure and semantics of clinical documents. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. The document can be sent inside an HL7 message and can exist independently, outside a transferring message. The first release of the standard has attempted to fill an important gap by addressing common and largely narrative clinical notes. It deliberately leaves out certain advanced and complex semantics, both to foster broad implementation and to give time for these complex semantics to be fleshed out within HL7.

Being a part of the emerging HL7 version 3 family of standards, the CDA derives its semantic content from the shared HL7 Reference Information Model and is implemented in Extensible Markup Language. The HL7 mission is to develop standards that enable semantic interoperability across all platforms. The HL7 version 3 family of standards, including the CDA, are moving us closer to the realization of this vision.

■ J Am Med Inform Assoc. 2001;8:552–569.

Affiliations of the authors: Kaiser Permanente, La Palma, California (RHD, PVB, JEM); alschuler.spinosa, East Thetford, Vermont (LA); Mayo Clinic, Rochester, Minnesota (CB); University of Southern California, Los Angeles (SLB, DE); Isogen International, Dallas, Texas (EK); University of Southern California, Santa Monica (TL).

Correspondence and reprints: Robert H. Dolin, MD, Kaiser Permanente, 5 Centerpointe Drive, La Palma, California 90623; e-mail: <Robert.H.Dolin@kp.org>.

Received for publication: 1/2/01; accepted for publication: 5/23/01.

Health Level 7 (HL7)¹ is a standards-setting organization accredited by the American National Standards Institute (ANSI). They have developed communication protocols widely used in the United States, with growing international recognition and implementations. A vendor- and provider-supported organization, its mission is to provide standards for the exchange, management, and integration of data that support clinical patient care and the management, delivery, and evaluation of health care servic-

es. This encompasses the complete life cycle of a standards specification—development, adoption, market recognition, utilization, and adherence. The HL7 specifications are unified by shared reference models of the health care and technical domains. The HL7 version 2.4 messaging standard is currently in use, and version 3, which represents several fundamental changes to the HL7 messaging approach, is in an advanced stage of development.^{2,3}

Many people know of HL7 as an organization that creates health care messaging standards. Health Level 7 is also developing standards for the representation of clinical documents (such as discharge summaries and progress notes). These document standards make up the HL7 Clinical Document Architecture (CDA). The HL7 Clinical Document Architecture, release 1.0, became an ANSI-approved HL7 Standard in November 2000.⁴

This article is intended to serve as an introduction to the CDA standard. It is geared toward medical informaticians who do not have significant familiarity with HL7 version 3, and it is intended to introduce the approach and objectives used in the creation of the standard and present an overview of the standard—not sufficient for implementation but sufficiently detailed to enable the reader to understand the scope and contents of the standard. Interested readers looking for detailed descriptions are encouraged to contact HL7 (www.hl7.org) for a copy of the normative specification.

Overview of the CDA

The need for a clinical document standard stems from the desire to unlock the considerable clinical content currently stored in free-text clinical notes and to enable comparison of content from documents created on information systems of widely varying characteristics. Given the variability in clinical notes, including structure, underlying information models, degree of semantic encoding, use of standard health-care terminologies, and platform- and vendor-specific features, it is currently difficult to store and exchange documents with retention of standardized semantics over both time and distance. And while the current CDA standard does not fully enable these main driving objectives, it takes us a step closer.

The CDA is a document markup standard that specifies the structure and semantics of “clinical documents.” A clinical document is a documentation of observations and services and has the following defining characteristics:

- *Persistence.* A clinical document continues to exist in an unaltered state, for a time period defined by local and regulatory requirements.
- *Stewardship.* A clinical document is maintained by a person or organization entrusted with its care.
- *Potential for authentication.* A clinical document is an assemblage of information that is intended to be legally authenticated.
- *Wholeness.* Authentication of a clinical document applies to the whole and does not apply to portions of the document without the full context of the document.
- *Human readability.* A clinical document is human readable.

Many draft and existing standards have helped inform the development of the CDA,^{5–14} and several guiding principles have driven the design:

- *Give priority to documents generated by clinicians involved in direct patient care.* There are many requirements and uses for clinical information, such as direct patient care, outcome research, and public health reporting. The CDA will give priority to defining documents that are created by clinicians involved in direct patient care, assuming the other uses will be derivable. The CDA will define documents produced by providers seeing patients and will not define views or downstream uses of those documents.
- *Minimize the technical barriers needed to implement the Standard.* There are estimated to be hundreds of thousands of nonstandardized clinical documents in existence. The CDA will facilitate standardization of these documents by allowing cost effective implementation across as wide a spectrum of systems as possible; by supporting exchange of human-readable documents between users, including those with different levels of technical sophistication; by enabling a wide range of post-exchange processing applications; by providing compatibility with a wide range of document creation applications; and by using non-health-care-specific standards where possible.
- *Promote longevity of all information encoded according to this architecture.* The CDA documents will be application- and platform-independent and can be viewed and edited by a number of tools, both now and in the future.
- *Promote exchange that is independent of the underlying transfer or storage mechanism.* The ability to

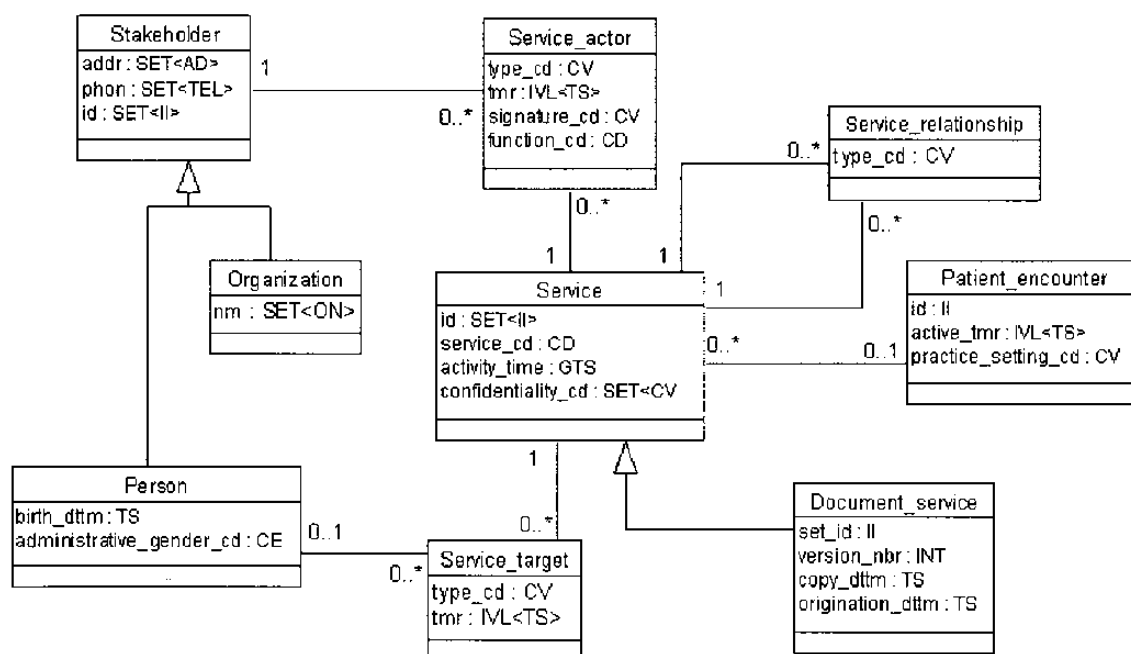


Figure 1 Subset of RIM version 0.98. A small subset of the HL7 Referenced Information Model, Version 0.98, used in the derivation of the CDA.

exchange or store CDA documents will be application- and platform-independent. These documents can be exchanged in HL7 messages, via e-mail, on a floppy disc, etc. A CDA document can be stored as an independent file, within a document management system, within a database, etc.

- *Enable policy makers to control their own information requirements without extension to this specification.* The CDA will define an extensibility mechanism that allows local implementations to represent information that is not formally represented in the standard.

The CDA is part of the HL7 version 3 family of standards. This family, which includes both CDA and the evolving version 3 message standards, all derive their semantic content from the shared HL7 Reference Information Model (RIM)¹⁵ and are implemented in Extensible Markup Language (XML).¹⁶ The process for generating an XML-based implementation from the RIM is part of the version 3 development methodology.^{2,17} The exact style of HL7's XML representation was a carefully considered balance of technical, practical, and functional considerations.^{18–22}

At the heart of the HL7 version 3 family of standards lies a shared information model, the HL7 RIM.¹⁴ An object-oriented model created as part of the version 3 methodology, the RIM is a large pictorial representation of clinical data and identifies the life cycles of events that messages and documents convey. All ver-

sion 3 products derive their semantic content from the RIM. The various committees in HL7 engage in an iterative consensus process to continually refine the RIM to meet identified use cases. Figure 1 is a small subset of RIM version 0.98 that was used in the derivation of the CDA.

The RIM includes a new set of data types developed for use in the HL7 version 3 family of standards.²³ These data types include some of the familiar ones used in HL7 v2.x messaging, such as STRING (ST), INTEGER (INT), and TIME STAMP (TS), as well as new data types such as ENCODED DATA (ED), which supports multimedia; INTERVAL of TIME (IVL<TS>), which allows for the expression of a time range; and CONCEPT DESCRIPTOR (CD), which supports the post-coordination of codes (or, stated in another way, the combining of codes from a terminology to create a new concept).

A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. The document can be sent inside an HL7 message and can exist independently outside a transferring message. There is a critical interdependence between clinical documents and document management systems. If CDA documents are viewed outside the context of a document management system, it cannot be known with certainty whether or not the viewed document is current.

Clinical documents can be revised, and they can be appended to existing documents. In practice, it is impossible to guarantee an explicit forward pointer from an outdated version to the newer version. Document management systems and HL7 messages that carry CDA documents convey critical contextual information that minimizes the risk of viewing superseded information.

The "A" in "CDA"

The complete CDA will include a hierarchic set of document specifications. This hierarchy is referred to as an "architecture" (the "A" in "CDA"). This architecture, which can be thought of as a set of hierarchically related XML Document Type Definitions (DTD) or schemas,^{24,25} is envisioned for future releases of the CDA standard. The current CDA standard has defined only the top node of the hierarchy, known as "CDA Level One."

Level One is sufficiently detailed to represent largely narrative clinical notes. The specification defines the document header in detail, while the document body is largely structural. Level One is intended to minimize the technical barriers to adoption of the standard while providing a gentle introduction to the RIM. It intentionally lacks some of the complex semantics that will be used in HL7 version 3 messaging (such as the ability to fully encode orders and observations). It is hoped that the provision of deeper levels of the architecture will provide a migration pathway for implementers to iteratively add greater markup to clinical documents.

Level Two is envisioned as a set of templates or constraints that can be layered on top of the CDA Level One specification. A template may, for instance, specify that a document of type "history-and-physical" will contain a mandatory "subjective" section; a mandatory "physical-examination" section containing a mandatory "vital-signs" section and an optional "cardiovascular-examination" section; and a mandatory "assessment" section followed by a "plan" section. This type of template development will require considerable domain knowledge and participation from professional societies so that created templates are widely embraced and supported. Health Level 7 is now in the process of establishing necessary liaison relationships with professional societies.

Level Three of the CDA will add additional RIM-derived markup to the CDA Level One specification, enabling clinical content to be formally expressed to the extent that is modeled in the RIM or to the extent that it can be expressed in an HL7 version 3 message.

This could, for instance, allow an order message to be extracted from a clinical document, allow for a detailed representation of symptoms and findings, and allow for billing codes to be automatically extracted.

The clinical content of CDA documents will remain invariable across all levels of the architecture. Each level both enables and enhances the standardized expression of richer shared semantics. Thus, a single report can be marked up as a Level One, Level Two, or Level Three document, and its clinical content will not vary. What will vary between the levels are the degree to which clinical content can be machine processible in an exchange context and the degree to which clinical document specifications can impose constraints on content.

Clinical Document Architecture Technical Overview

A CDA document has a header and a body. The header conveys the context in which the document was created, and the body contains the informational (factual) statements that make up the actual content of the document. The purpose of the header is to make clinical document exchange possible across and within institutions; facilitate clinical document management; and facilitate compilation of an individual patient's clinical documents into a lifetime electronic health record.

The header has four logical components:

- *Document information* identifies the document, defines confidentiality status, and describes relationships to other documents and orders.
- *Encounter data* describes the setting in which a documented encounter occurred
- *Service actors* include those who authenticate the document, those intended to receive a copy of the document, document originators and transcriptionists, and health care providers who participated in the service(s) being documented
- *Service targets* include the patient and other significant participants (such as family members).

The CDA body comprises sections, paragraphs, lists, and tables. These structural components have captions, can nest (so that, for instance, a section can contain a section), and can contain character data, multimedia, and codes drawn from standard terminologies.

Appendix 1 shows a sample clinical document, and Appendix 2 shows a CDA-encoding of that sample document. The sample CDA document includes

many optional components (including some that are not explicitly shown in Appendix 1) to better illustrate more features. Line numbers in Appendix 2 are be referenced throughout this technical overview to help illustrate the described concepts.

CDA Header (Appendix 2, Lines 4–120)

Document Information (Appendix 2, Lines 5–25)

A CDA document can be an original document, an addendum to an existing document, or a revision of an existing document. An original document is the first version of a document. An addendum is an appendage to an existing report that contains supplemental information. The parent report being appended remains in place and its content and status are unaltered. A replacement report is a revision that replaces an existing report. The replacement report gets a new globally unique <id>* value but uses the same value for <set_id> as the parent report being replaced, and increments the value of <version_nbr> by 1 (Appendix 2, lines 5–7). The parent report is considered superseded but is still retained in the system for historical reference. Addendum and replacement documents reference the parent report via a <document_relationship> (Appendix 2, lines 13–20).

Documents can also be related to one or more orders. The <fulfills_order> component relates the current document to the unique identifier(s) of the orders that have been fulfilled (Appendix 2, lines 21–25) by this document.

Every document has a required document type code, <document_type_cd> (Appendix 2, lines 8–9), which classifies the document. The externally defined vocabulary domain for <document_type_cd> is drawn from LOINC (Logical Observation Identifiers, Names and Codes).^{26,27} The creation of an ontology of clinical document names and codes in LOINC was accelerated by the contributions of the Document Ontology Task Force, a consortium of standards and vocabulary organizations sponsored by HL7 with support from the Veterans Health Administration, charged with this task. On the basis of the analysis of more than 2,000 clinical document names, the task force formulated a terminology model that was given to LOINC to guide the creation of fully specified document names.^{28,29}

Many temporal events come into play in the creation and validation of a CDA document. Some of the relevant times are part of the document information

(e.g., <origination_dttm>, which represents the time a document is initiated), while other times relate to the encounter (e.g., <encounter_tmtr>), and the time during which various service actors and targets played a role (e.g., <participation_tmtr>). Some temporal events can be represented as a specific point in time (indicated by an XML element name ending in “_dttm” and using the HL7 “point in time” [TS] data type), while other temporal events include time points or time intervals (indicated by an XML element name ending in “_tmtr” and using the HL7 “interval of time” [IVL<TS>] or “general time specification” [GTS] data type).

The CDA header indicates document confidentiality status using the coded <confidentiality_cd> component (Appendix 2, lines 11–12).³⁰ A single document confidentiality status indicator in the header will apply to the entire document. Alternatively, to assign different levels of confidentiality to different portions of the document, implementers can include more than one confidentiality status indicator in the header and then reference the appropriate indicator from a structure within the document body.

Encounter Data (Appendix 2, Lines 26–38)

Encounter data include an encounter identifier, an encounter time stamp, and a location. An encounter also includes an optional practice setting code, <practice_setting_cd>, which is a categorization of the clinical setting (e.g., cardiology clinic, primary care clinic, rehabilitation hospital, skilled nursing facility) in which care is delivered (Appendix 2, lines 28–29).

Service Actors (Appendix 2, Lines 39–84)

All people and organizations concerned with a CDA document can be associated with the document as either service actors (described here) or service targets (described next). Service actors include those who authenticate the document, those intended to receive a copy of the document, document originators and transcriptionists, and health care providers who participated in the services being documented. Service actors are capable of and accountable for their independent decisions.

Specific service actors defined in the CDA header include <authenticator>, <legal_authenticator> (Appendix 2, lines 39–57), <originator> (Appendix 2, lines 58–70), <intended_recipient>, <originating_organization> (Appendix 2, lines 71–77), <transcriptionist>, and <provider> (Appendix 2, lines 78–84).

The CDA is a standard that specifies the structure of exchanged clinical documents. In a case in which a

* In this text, XML element names that are in the CDA standard are surrounded by angle brackets (< >).

local document is transformed into a CDA document for exchange, authentication occurs on the local document and that fact is reflected in the exchanged CDA document. A CDA document can reflect the unauthenticated, authenticated, or legally authenticated state. The unauthenticated state exists when no authentication information has been recorded (i.e., it is the absence of being either authenticated or legally authenticated). An authenticated document exists when one or more providers have attested to the accuracy of the document. A legally authenticated document exists when the person who is legally responsible for the document has attested to its accuracy. Both authentication and legal authentication require that a document be signed manually or electronically by the responsible person. Although electronic signatures are not currently part of the CDA header, the CDA header does require the acquisition of a signature to be documented via the <signature_cd> component (Appendix 2, line 42).

The CDA documents can be originated (authored) by human beings or machines, or both. The <originator> element is used to indicate human origination (Appendix 2, lines 58–70), and the <originating_device> element is used to indicate machine origination (Appendix 2, lines 105–16). The CDA header requires the specification of one or more document originators but does not require that there be a human originator.

Intended recipients, indicated by <intended_recipient>, are those people to whom a copy of the document is to be sent. The CDA header can specify the organization from which the document originates and that is in charge of maintaining the document using the <originating_organization> component (Appendix 2, lines 71–77). The CDA header requires the specification of one or more health care providers, indicated by <provider>, who participated in the services being documented (Appendix 2, lines 78–84).

Service Targets (Appendix 2, Line 85–116)

Service targets are physical entities, including living subjects and inanimate material, that are typically the object of services being documented. Service targets include the patient and other significant participants (such as family members).

The CDA header requires one and only one value for <patient> (Appendix 2, lines 85–104). The value of the <patient> component indicates whose medical record this document belongs to. By default, the <patient> is also the principal subject of the services being documented. The <patient> is defined as hav-

```
<!--ATTLIST CDA_Body_Structural_Elements
...
confidentiality IDREFS #IMPLIED
originator IDREFS #IMPLIED
xml:lang NMTOKEN #IMPLIED
...
-->
```

Figure 2 Structure “context” in CDA body. Attribution ascribed to structural components in the CDA Body (sections, paragraphs, lists, tables) define the structure’s “context”. These attributes are applicable to the contents of the structure, and are inherited by nested structures, unless overridden.

ing the same components as any other person, plus a date of birth, <birth_dttm> (Appendix 2, line 102), and a gender, <administrative_gender_cd> (Appendix 2, line 103).

In addition to <patient>, other people (e.g., a patient’s child, parent, beneficiary) may play a role in certain circumstances or with certain types of documents. The CDA header can specify one or more of these people via the <service_target> component.

CDA Body (Appendix 2, Line 121–268)

Document Structures

The CDA body comprises “structures” that include sections, paragraphs, lists, and tables. These structures have captions, can nest (so that, for instance, a section can contain a section), and can contain “entries” such as character data, multimedia, and codes drawn from standard terminologies.

All CDA structures have a “context,” meaning that attribution assigned to a structure is applicable to the contents contained in that structure, and are inherited by nested structures unless overridden. Structure attributes of a CDA include confidentiality, origination, and human language (Figure 2).

The confidentiality attribute and the originator attribute reference values are expressed in the CDA header (using the XML “IDREFS” data type). The confidentiality attribute references the <confidentiality_cd> component (Appendix 2, line 121), whereas the originator attribute references the <originator> or the <originating_device> component (Appendix 2, line 227). Thus, all applicable confidentiality and origina-

tor values are expressed in the CDA header and referenced by structures in the body. The language attribute is used to indicate the human language of character data and is specified using the "xml:lang" attribute as defined in the XML 1.0 Recommendation:

A special attribute named xml:lang may be inserted in documents to specify the language used in the contents and attribute values of any element in an XML document. In valid documents, this attribute, like any other, must be declared if it is used. The values of the attribute are language identifiers as defined by the IETF (Internet Engineering Task Force) RFC 1766: *Tags for the Identification of Languages*, ed. H. Alvestrand, 1995.

A CDA structure has a caption, <caption> (Appendix 2, lines 123–126), which is an optionally coded heading or label for a container. The vocabulary domain used to codify a caption (Appendix 2, line 124) is externally defined by LOINC. The creation of section codes in LOINC was facilitated by input from the Document Ontology Task Force.^{10,28,29,31}

The CDA <list> is patterned after the HTML list structure and contains one or more <item> elements (Appendix 2, lines 138–142). The list_type attribute specifies whether the <list> is ordered or unordered, with unordered being the default. An ordered list is used when the ordering of list items is meaningful. The CDA <table> is a modification of the XHTML⁹ table model, which has removed some of the formatting tags. In the CDA body, any information can be presented as a table. The table markup is for presentation purposes only and, unlike a database table, does not possess meaningful field names.

In general, CDA structures are similar to HTML structures. Because of the similarity, transforming CDA structures into HTML structures for the sake of viewing a CDA document on a Web browser is very straightforward, and CDA documents can be authored using tools that generate HTML. The resulting HTML can be transformed into CDA. It should be noted that HTML will not support much of the detailed markup in the CDA, but it is expected to retain compatibility with the structural components even as the CDA standard evolves.

Document Entries

Document entries in a CDA occur in structures and include character data; coded entries, <coded_entry>; a recursively nesting wrapper, <content>; a generic referencing mechanism, <link>; and multimedia, <observation_media>.

The CDA element <coded_entry> (Appendix 2, lines 239–242) is used to insert codes from HL7-recognized

coding schemes into CDA documents. The <content> element (Appendix 2, lines 236–243) can be thought of as a fine-grained wrapper and explicit anchor. Because it can nest recursively, the <content> element enables wrapping a string of plain text down to as small a chunk as desired. Coded entries can reference these <content> anchors to indicate the original text that supports the use of the code, in a manner similar to that described by Friedman et al.³² The process is shown in Figure 3. In the first case, <coded_entry> is used to simply insert a code into a CDA document, without referencing the original text. In the second case, <coded_entry> explicitly references the original text that supports the assigned code.

The primary intent of <coded_entry> is to facilitate document indexing, search, and retrieval and to pro-

Case 1

<content>

Asthma, with prior smoking history. Difficulty weaning off steroids.

Will try gradual taper.

<coded_entry>

<coded_entry.value

V="D2-51000" S="SNOMED" DN="Asthma"/>

</coded_entry>

</content>

Case 2

<content>

<content ID="String001">Asthma</content>, with prior smoking history.

Difficulty weaning off steroids. Will try gradual taper.

<coded_entry>

<coded_entry.value ORIGTXT="String001"

V="D2-51000" S="SNOMED" DN="Asthma"/>

</coded_entry>

</content>

Figure 3 Referencing original text with <coded_entry>. In the first case, <coded_entry> is used to simply insert a code into a CDA document, without referencing the original text. In the second case, <coded_entry> explicitly references the original text that supports the assigned code.

vide a standard convention for insertion of locally meaningful codes. The CDA body lacks a robust semantic for the full representation of a code's context. Thus, when a computer encounters a code for "chest pain" occurring in a "history of present illness" section, the computer cannot reliably determine whether "chest pain" refers to the patient or to the patient's family member, nor can it reliably determine whether "chest pain" is present or stated to be absent. For now, these determinations rely on human interpretation. As consensus on these issues develops within HL7, the shared understanding will be reflected in the RIM, and they will be incorporated into CDA.

Modifiers of coded entries (e.g., Mass :: Has-color :: Blue; Mass :: Has-shape :: Round) can be formally expressed using the power of the coded entry's Concept Descriptor (CD) data type, which allows for post-coordination, or the combining of codes from a terminology to create a new concept. The use of post-coordination raises some interesting challenges in data extraction, since many concepts that exist in the underlying terminology (such as "nodule-of-skin") can also be constructed via post-coordination ("lesion-of-skin :: Has-morphology :: nodule").^{33,34} While the approach to this challenge is outside the scope of the current paper, it is of active and ongoing interest to HL7.

The CDA <link> is a generic referencing mechanism, similar to the HTML anchor tag. Several groups (such as the World Wide Web Consortium) are actively developing formal link specifications.³⁵ When a suitable open standard is available and implemented, it will be reviewed with the intent to incorporate it into the CDA specification. The <observation_media> element (Appendix 2, lines 194–198) represents media that is logically a part of a CDA document but is stored outside the document and incorporated by reference, in a manner similar to the HTML image tag.

Localization and Transformation issues

Those implementing structured document systems for health care may choose to adopt CDA directly into their application or environment or may implement their own internal document representation standards. In the first case, local implementations of CDA will often require the ability to extend or localize the shared semantics expressed by CDA. In the second case, exchange of CDA documents will require that local document instances be transformed into CDA-compliant instances.

The implementation of localization in CDA parallels the "version compatibility definition" in HL7 V2.x that enables new fields, segments, and components to be introduced at the tail end of the normative specification. In the process of creating the CDA XML DTD, the CDA specification adds a <local_header> element to the end of every XML content model in the header (Appendix 2, lines 117–119) and a <local_markup> element to the end of every XML content model in the body. These localization tags are optionally repeating and recursive. The "descriptor" attribute describes the element, and the value can be drawn from a local vocabulary domain. The "ignore" attribute tells the receiver to ignore just the local tag (ignore="markup") or to ignore the tag and all contained content (ignore="all").

Transformation of locally defined document instances into CDA-compliant instances can be performed with special-purpose SGML^{36,37} and XML transformation languages (such as Hytime Architectural Forms,¹² Document Style Semantics and Specification Language,³⁸ and Extensible Stylesheet Language³⁹) or with general programming languages. Where possible, local semantics should be mapped into the CDA's shared semantics. In the case where local semantics have not been standardized in CDA, mapping to CDA's <local_header> or <local_markup> is appropriate.

Each transformation language has potential syntactic limitations in the type of mappings that can be formally expressed. In addition, there are semantic mappings that cannot be resolved regardless of the transformation language employed. Although the discussion here focuses on mapping from one DTD to another, the issues are no different from those that arise in the more general sense of mapping from one information model to another. Elements in the local instance may be in different order than in CDA (e.g., <dob> <name> vs. <person_name> <birth_dttm>). The source may be missing a required element (e.g., CDA requires a globally unique document identifier). The source and CDA may have different vocabulary domains (e.g., "male" vs. "M"). Elements may have different data types (e.g., "January 13, 1923" vs. "1923-01-13"). The source may have coarser granularity (i.e., less detailed markup) than CDA (e.g., <name> Henry Levin, the 7th </name> vs. <person_name> <GIV V="Henry"/> <FAM V="Levin"/> <SFX V="the 7th"/> </person_name>).

It is likely that transformation issues will be minimized to the extent that the CDA specifications and the underlying RIM are unambiguous. Where local

requirements not met by the CDA are identified, they should be considered candidate proposals to enhance HL7 or CDA semantics.

Conclusion

The first release of the ANSI-approved HL7 CDA represents the culmination of close to four years of effort. With this standard, HL7 has entered the realm of defining the structure and semantics of clinical documents. This first standard has attempted to fill a huge gap by standardizing common clinical notes such as history and physicals, discharge summaries, and progress notes. It deliberately leaves out certain advanced and complex semantics both to foster broad implementation and to give time for these complex semantics to be flushed out in the RIM. Health Level 7 is currently balloting the new version 3 messaging specifications; as a result, considerable energy is going into refining the RIM. It is expected that future releases of CDA will include deeper layers of the architecture that encode richer semantics and that the current standard will serve as a stepping stone, allowing users to implement these releases progressively.

More and more, decision support applications will communicate with electronic health records via an HL7 interface. For this transfer of information to be rich, detailed, and unambiguous, a high degree of semantic interoperability between applications is needed. The HL7 version 3 family of standards, including the CDA, are moving us closer to the realization of this vision.

The history of the CDA can be traced to November 1996, when John Mattison and John Spinoso developed a plan to bring together leaders with expertise in both document markup and clinical informatics and develop a standard approach to creating and communicating clinical documents. After a series of meetings, a subgroup of the project launched "Operation Jumpstart," and a group of enthusiastic participants (Liora Alschuler, Ron Capwell, Robert Dolin, Daniel Essin, Jasen Fici, Lloyd Harding, Eliot Kimber, Anil Sethi, Rachael Sokolowski, John Spinoso, Michael Toback, and Jason Williams) met at the Kona Mansion on Lake Winnepesaukee, New Hampshire the week of July 7, 1997, to draft the "Kona Proposal," which over time would morph into the CDA.

Since then, the CDA could never have evolved into a national standard without the collaborative spirit and the enormous personal contributions of Carl Adler, Woody Beeler, Fred Behlen, Dean Bidgood, Carol Broverman, Hans Buitendijk, John Carter, Michal Coleman, Don Connelly, Gary Dickinson, Steve Doubleday, Joachim Dudeck, Al Figler, Leo Fogarty, Gerard Freriks, Ed Hammond, Richard Harding, Michael Henderson, Stan Huff, Juggy Jagannathan, Ed Jones, Tom Lincoln, John Majerus, David Markwell, John Mattison, Bob Moe, Wes Rishel, Angelo Rossi-Mori, Gunther Shadow, Anne Shanney, Wayne Tracy, and Chris Zingo.

The authors acknowledge the contributions and domain expertise of the HL7 Medical Records/Information Management Technical Committee and the Orders and Observation Technical Committee. The close working relationship they have had with these groups ensures the consistency of the various work products generated by HL7.

They also thank Jonathan Lukoff, MD, William W. Stead, MD, and the JAMIA reviewers for their critique of the manuscript.

References ■

1. Health Level 7. HL7 Web site. Available at: <http://www.hl7.org>. Accessed Oct 12, 2001.
2. Beeler GW Jr. Taking HL7 to the next level. *MD Comput*. 1999;16(2):21-4.
3. Dolin RH. Advances in data exchange for the clinical laboratory. *Clin Lab Med*. 1999;19(2):385-419.
4. Alschuler L, Dolin RH, Boyer S, Beebe C (eds). HL7 Clinical Document Architecture Framework, Release 1.0. ANSI-approved HL7 Standard, Nov 2000. Ann Arbor, Mich.: Health Level Seven, Inc., 2000.
5. Dolin RH, Alschuler L, Boyer S, Beebe C. An update on HL7's XML-based document representation standards. *Proc AMIA Annu Symp*. 2000:190-4.
6. Dolin RH, Alschuler L, Behlen F, et al. HL7 document patient record architecture: an XML document architecture based on a shared information model. *AMIA Annu Fall Symp*. 1999:52-6.
7. Lincoln TL, Essin DJ, Anderson R, Ware WH. The Introduction of a New Document Processing Paradigm into Health Care Computing. A National Institute of Standards and Technology Center for Advanced Information Technology White Paper. Gaithersburg, Md.: NIST, 1994.
8. Essin DJ, Lincoln TL. An information model for medical events. *Proc Annu Symp Comput Appl Med Care*. 1994:509-13.
9. Essin DJ. Intelligent processing of loosely structured documents as a strategy for organizing electronic health care records. *Methods Inf Med*. 1993;32:265-68.
10. World Wide Web Consortium (W3C). XHTML 1.0: The Extensible HyperText Markup Language—A Reformulation of HTML 4 in XML 1.0. Recommendation, Jan 26, 2000. Available at: <http://www.w3.org/TR/xhtml1/>. Accessed Oct 12, 2001.
11. European Committee for Standardization, Technical Committee for Health Informatics (CEN TC251). Four-part EHCR Message Standard. Brussels, Belgium: CEN TC251. Publication CEN ENV 13606.
12. National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM), Supplement 23: Structured Reporting. Rosslyn, Va: NEMA, Oct 29, 1999.
13. International Organization for Standardization. Architectural Forms Definition Requirements (AFDR). Geneva, Switzerland: ISO, 1990. Publication ISO/IEC 10744.
14. Freed N, Borenstein N. Multipurpose Internet Mail Extensions (MIME) Part 2: Media Types. Request for comments (RFC) 2046. Internet Engineering Task Force Web site. Available at: <http://www.ietf.org/rfc.html>. Accessed Oct 12, 2001.
15. Health Level Seven, Inc. HL7 Reference Information Model. Ann Arbor, Mich: Health Level Seven, 1994. Available at: http://www.hl7.org/library/data-model/RIM/model-page_non.htm. Accessed Oct 12, 2001.
16. World Wide Web Consortium (W3C). Extensible Markup Language (XML) 1.0. Recommendation, Feb 10, 1998. Available at: <http://www.w3.org/TR/1998/REC-xml-19980210.html>. Accessed Oct 12, 2001.
17. Beeler GW Jr. HL7 version 3: an object-oriented methodology for collaborative standards development. *Int J Med Inform*. 1998;48(1-3):151-61.
18. Dolin RH, Biron PV (eds). Using XML as a Supplementary Messaging Syntax for HL7 Version 2.3.1. HL7 Recommendation, Nov 1999. Ann Arbor, Mich.: Health Level Seven, Inc., 1999.
19. Dolin RH, Rishel W, Biron PV, Spinoso J, Mattison JE. SGML

- and XML as interchange formats for HL7 messages. *Proc AMIA Symp.* 1998:720-4.
20. Alschuler L, Dolin R, Spinosa J. SGML in healthcare information systems. *GCA SGML '97 Conference Proceedings*. Alexandria, Va.: Graphic Communications Association, 1997.
 21. Dolin RH, Alschuler L, Bray T, Mattison JE. SGML as a message interchange format in healthcare. *Proc AMIA Annu Fall Symp.* 1997:635-9.
 22. European Committee for Standardization, Technical Committee for Health Informatics (CEN TC251). Investigation of Syntaxes for Existing Interchange Formats to be Used in Healthcare. Brussels, Belgium: CEN TC251, Jan 1993.
 23. Bakken S, Campbell KE, Cimino JJ, Huff SM, Hammond WE. Toward vocabulary domain specifications for Health Level 7-coded data elements. *J Am Med Inform Assoc.* 2000;7(4):333-42.
 24. World Wide Web Consortium (W3C). XML Schema Part 1: Structures. W3C working draft, May 6, 1999. Available at: <http://www.w3.org/1999/05/06-xmlschema-1/>. Accessed Oct 12, 2001.
 25. World Wide Web Consortium (W3C). XML Schema Part 2: Datatypes. W3C working draft, May 6, 1999. Available at: <http://www.w3.org/1999/05/06-xmlschema-2/>. Accessed Oct 12, 2001.
 26. Logical Observations Identifiers, Names, and Codes (LOINC) Regenstrief Web site. Available at: <http://www.regenstrief.org/loinc/loinc.htm>. Accessed Oct 12, 2001.
 27. Huff SM, Rocha RA, McDonald CJ, et al. Development of the Logical Observations Identifiers, Names, and Codes (LOINC) vocabulary. *J Am Med Inform Assoc.* 1998;5:276-92.
 28. Document Ontology Task Force. Proposal for an Ontology for Exchange of Clinical Documents. Jul 19, 2000. HL7 Web site. Available at: <http://www.hl7.org/special/dotf/dotf.htm>. Accessed Oct 12, 2001.
 29. Frazier P, Rossi-Mori A, Dolin RH, Alschuler L, Huff SM. The creation of an ontology of clinical document names. In: Patel V, Rogers R, Haux R (eds). *MedInfo 2001. Studies in Health Technology and Informatics*, vol 84. Amsterdam, The Netherlands: IOS Press, 2001.
 30. Essin DJ, Lincoln TL: Healthcare information architecture: elements of a new paradigm. Presented at: National Computer Security Conference; Oct 1994; Baltimore, Maryland.
 31. National Health Service (U.K.) Information Authority. Towards an Information Standard for Organizing Clinical Communications. Position paper, final draft. Birmingham, UK: NHSIA, Mar 2000.
 32. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. *J Am Med Inform Assoc.* 1999;6(1):76-87.
 33. Brown PJB, Sonksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. *J Am Med Inform Assoc.* 2000;7(4):392-403.
 34. Rector AL. The interface between information, terminology, and inference models. *MedInfo*. 2001. In: Patel V, Rogers R, Haux R (eds). *MedInfo 2001. Studies in Health Technology and Informatics*, vol 84. Amsterdam, The Netherlands: IOS Press, 2001.
 35. World Wide Web Consortium (W3C). XML Linking Language (XLink). Proposed recommendation, Dec 12, 2000. Available at: <http://www.w3.org/TR/xlink/>. Accessed Oct 12, 2001.
 36. International Organization for Standardization. Information processing—Text and office systems—Standard Generalized Markup Language (SGML). Geneva, Switzerland: ISO, Oct 1986. Publication ISO 8879:1986.
 37. International Organization for Standardization. Information processing—Text and office systems—Standard Generalized Markup Language (SGML), Amendment 1. Geneva, Switzerland: ISO Jul 1988. Publication ISO 8879:1986/Amendment 1.
 38. International Organization for Standardization. Document Style Semantics and Specification Language (DSSSL). Geneva, Switzerland: ISO, 1996. Publication ISO/IEC 10179:1996.
 39. World Wide Web Consortium (W3C). The Extensible Stylesheet Language (XSL). W3C Web site. Available at: <http://www.w3.org/Style/XSL/>. Accessed Oct 12, 2001.

Appendix A
SAMPLE DOCUMENT

Good Health Clinic Consultation note

Consultant: Robert Dolin, MD
Date: April 7, 2000
Patient: Henry Levin, the 7th **MRN:** 12345 **Sex:** Male
Birthdate: September 24, 1932

History of Present Illness

Henry Levin, the 7th is a 67 year old male referred for further asthma management. Onset of asthma in his teens. He was hospitalized twice last year, and already twice this year. He has not been able to be weaned off steroids for the past several months.

Past Medical History

- Asthma
- Hypertension
- Osteoarthritis, right knee

Medications

- Theodur 200mg BID
- Proventil inhaler 2puffs QID PRN
- Prednisone 20mg qd
- HCTZ 25mg qd

Allergies

- Penicillin - Hives
- Aspirin - Wheezing

Social History

- Smoking :: 1 PPD between the ages of 20 and 55, and then he quit.
- Alcohol :: Rare

Physical Exam

- **Vital Signs** :: BP 118/78; Wt 185lb; Resp 16 and unlabored; T 98.6F; HR 86 and regular.
- **Skin** :: Erythematous rash, palmar surface, left index finger.



- **Lungs** :: Clear with no wheeze. Good air flow.
- **Cardiac** :: RRR with no murmur, no S3, no S4.

Labs

- CXR 02/03/1999: Hyperinflated. Normal cardiac silhouette, clear lungs.
- Peak Flow today: 260 l/m.

Assessment

- Asthma, with prior smoking history. Difficulty weaning off steroids. Will try gradual taper.
- Hypertension, well-controlled.
- Contact dermatitis on finger.

Plan

- Complete PFTs with lung volumes.
- Chem-7
- Provide educational material on inhaler usage and peak flow self-monitoring.
- Decrease prednisone to 20qOD alternating with 18qOD.
- Hydrocortisone cream to finger BID.
- RTC 1 week.

Signed by: Robert Dolin, MD April 8, 2000

Appendix B

SAMPLE CDA DOCUMENT

This CDA document uses rich markup to illustrate many features of CDA. Many of the elements included in this sample are optional. This is a validated CDA document that conforms to the Standard.

```

1  <?xml version="1.0"?>
2  <!DOCTYPE levelone PUBLIC "-//HL7//DTD CDA Level One 1.0//EN" "levelone_1.0.dtd">
3  <levelone>
4    <clinical_document_header>
5      <id EX="a123" RT="2.16.840.1.113883.3.933"/>
6      <set_id EX="B" RT="2.16.840.1.113883.3.933"/>
7      <version_nbr V="2"/>
8      <document_type_cd V="11488-4" S="LOINC"
9        DN="Consultation note"/>
10     <origination_dttm V="2000-04-07"/>
11     <confidentiality_cd ID="CONF1" V="N" S="HL7_Confidentiality"/>
12     <confidentiality_cd ID="CONF2" V="R" S="HL7_Confidentiality"/>
13     <document_relationship>
14       <document_relationship.type_cd V="RPLC"/>
15       <related_document>
16         <id EX="a234" RT="2.16.840.1.113883.3.933"/>
17         <set_id EX="B" RT="2.16.840.1.113883.3.933"/>
18         <version_nbr V="1"/>
19       </related_document>
20     </document_relationship>
21     <fulfills_order>
22       <fulfills_order.type_cd V="FLFS"/>
23       <order><id EX="x23ABC" RT="2.16.840.1.113883.3.933"/></order>
24       <order><id EX="x42CDE" RT="2.16.840.1.113883.3.933"/></order>
25     </fulfills_order>
26     <patient_encounter>
27       <id EX="KPENC1332" RT="2.16.840.1.113883.3.933"/>
28       <practice_setting_cd V="GIM"
29         S="HL7_PracticeSetting" DN="General internal medicine clinic"/>
30       <encounter_tmr V="2000-04-07"/>
31       <service_location>
32         <id EX="KOLPa123" RT="2.16.840.1.113883.3.933"/>
33         <addr>
34           <HNR V="970"/><STR V="Post St"/><DIR V="NE"/>
35           <CTY V="Alameda"/><STA V="CA"/><ZIP V="94501"/>
36         </addr>
37       </service_location>
38     </patient_encounter>
39     <legal_authenticator>
40       <legal_authenticator.type_cd V="SPV"/>
41       <participation_tmr V="2000-04-08"/>
42       <signature_cd V="S"/>
43       <person>
44         <id EX="KP00017" RT="2.16.840.1.113883.3.933"/>
45         <person_name>
46           <nm>
47             <GIV V="Robert"/><FAM V="Dolin"/><SFX V="MD" QUAL="PT"/>
48           </nm>

```

```

49         <person_name.type_cd V="L" S="HL7_PersonNamePurpose"/>
50     </person_name>
51     <addr>
52         <HNR V="970"/><STR V="Post St"/><DIR V="NE"/>
53         <CTY V="Alameda"/><STA V="CA"/><ZIP V="94501"/>
54     </addr>
55     <telecom V="tel:(358)555-1234" USE="WFN"/>
56 </person>
57 </legal_authenticator>
58 <originator>
59     <originator.type_cd V="AUT"/>
60     <participation_tmr V="2000-04-07"/>
61     <person>
62         <id EX="KP00017" RT="2.16.840.1.113883.3.933"/>
63         <person_name>
64             <nm>
65                 <GIV V="Robert"/><FAM V="Dolin"/><SFX V="MD"/>
66             </nm>
67             <person_name.type_cd V="L" S="HL7_PersonNamePurpose"/>
68         </person_name>
69     </person>
70 </originator>
71 <originating_organization>
72     <originating_organization.type_cd V="CST"/>
73     <organization>
74         <id EX="M345" RT="2.16.840.1.113883.3.933"/>
75         <organization.nm V="Good Health Clinic"/>
76     </organization>
77 </originating_organization>
78 <provider>
79     <provider.type_cd V="CON"/>
80     <participation_tmr V="2000-04-07"/>
81     <person>
82         <id EX="KP00017" RT="2.16.840.1.113883.3.933"/>
83     </person>
84 </provider>
85 <patient>
86     <patient.type_cd V="PATSBJ"/>
87     <person>
88         <id EX="12345" RT="2.16.840.1.113883.3.933"/>
89         <person_name>
90             <nm>
91                 <GIV V="Henry"/><FAM V="Levin"/><SFX V="the 7th"/>
92             </nm>
93             <person_name.type_cd V="L" S="HL7_PersonNamePurpose"/>
94         </person_name>
95         <person_name>
96             <nm>

```

```

97         <GIV V="Hank"/><FAM V="Levin"/>
98     </nm>
99     <person_name.type_cd V="A" S="HL7_PersonNamePurpose"/>
100 </person_name>
101 </person>
102 <birth_dttm V="1932-09-24"/>
103 <administrative_gender_cd V="M" S="HL7_AdministrativeGender"/>
104 </patient>
105 <originating_device ID="DEV1">
106     <originating_device.type_cd V="ODV"/>
107     <device>
108         <id EX="devX3498" RT="2.16.840.1.113883.3.933"/>
109         <responsibility>
110             <responsibility.type_cd V="MNT" S="HL7_RelationshipResponsibility"/>
111             <person>
112                 <id EX="KP03257" RT="2.16.840.1.113883.3.933"/>
113             </person>
114         </responsibility>
115     </device>
116 </originating_device>
117 <local_header ignore="all" descriptor="MyLocalTag">
118     ... extra stuff that is only used locally ...
119 </local_header>
120 </clinical_document_header>
121 <body confidentiality="CONF1">
122     <section>
123         <caption>
124             <caption_cd V="10164-2" S="LOINC"/>
125             History of Present Illness
126         </caption>
127         <paragraph>
128             <content>
129                 Henry Levin, the 7th is a 67 year old male referred for further
130                 asthma management. Onset of asthma in his teens. He was hospitalized
131                 twice last year, and already twice this year. He has not been able to
132                 be weaned off steroids for the past several months.
133             </content>
134         </paragraph>
135     </section>
136     <section>
137         <caption>Past Medical History</caption>
138         <list>
139             <item><content>Asthma</content></item>
140             <item><content>Hypertension</content></item>
141             <item><content>Osteoarthritis, right knee</content></item>
142         </list>
143     </section>
144 </section>

```

```

145     <caption>
146         <caption_cd V="10160-0" S="LOINC"/>Medications
147     </caption>
148     <list>
149         <item><content>Theodur 200mg BID</content></item>
150         <item><content>Proventil inhaler 2puffs QID PRN</content></item>
151         <item><content>Prednisone 20mg qd</content></item>
152         <item><content>HCTZ 25mg qd</content></item>
153     </list>
154 </section>
155 <section>
156     <caption>Allergies</caption>
157     <list>
158         <item><content>Penicillin - Hives</content></item>
159         <item><content>Aspirin - Wheezing</content></item>
160     </list>
161 </section>
162 <section>
163     <caption>Social History</caption>
164     <list>
165         <item>
166             <content>
167                 Smoking :: 1 PPD between the ages of 20 and 55, and then he quit.
168             </content>
169         </item>
170         <item><content>Alcohol :: rare</content></item>
171     </list>
172 </section>
173 <section>
174     <caption>
175         <caption_cd V="11384-5" S="LOINC"/>Physical Examination
176     </caption>
177     <section>
178         <caption>
179             <caption_cd V="8716-3" S="LOINC"/>Vital Signs
180         </caption>
181         <list>
182             <item><content>BP 118/78</content></item>
183             <item><content>Resp 16 and unlabored</content></item>
184             <item><content>T 98.6F</content></item>
185             <item><content>HR 86 and regular</content></item>
186         </list>
187     </section>
188 <section>
189     <caption>
190         <caption_cd V="8709-8" S="LOINC"/>Skin
191     </caption>
192     <paragraph>

```



```

193         <content>Erythematous rash, palmar surface, left index finger.
194         <observation_media>
195             <observation_media.value MT="image/jpeg">
196                 <REF V="rash.jpeg"/>
197             </observation_media.value>
198         </observation_media>
199     </content>
200 </paragraph>
201 </section>
202 <section>
203     <caption>
204         <caption_cd V="11391-0" S="LOINC"/>Lungs
205     </caption>
206     <paragraph>
207         <content>Clear with no wheezes. Good air flow.</content>
208     </paragraph>
209 </section>
210 <section>
211     <caption>Cardiac</caption>
212     <paragraph>
213         <content>RRR with no murmur, no S3, no S4.</content>
214     </paragraph>
215 </section>
216 </section>
217 <section confidentiality="CONF2">
218     <caption>
219         <caption_cd V="11502-2" S="LOINC"/>Labs
220     </caption>
221     <list>
222         <item>
223             <content>
224                 CXR 02/03/1999: Hyperinflated. Normal cardiac silhouette, clear lungs.
225             </content>
226         </item>
227         <item originator="DEV1"><content>Peak Flow today: 260 l/m</content></item>
228     </list>
229 </section>
230 <section>
231     <caption>
232         <caption_cd V="11496-7" S="LOINC"/>Assessment
233     </caption>
234     <list>
235         <item>
236             <content>
237                 <content ID="String001">Asthma</content>, with prior smoking
238                 history. Difficulty weaning off steroids. Will try gradual taper.
239             <coded_entry>
240                 <coded_entry.value ORIGIN="String001"

```

```

241         V="D2-51000" S="SNOMED"/>
242     </coded_entry>
243 </content>
244 </item>
245 <item><content>Hypertension, well-controlled.</content></item>
246 <item><content>Contact dermatitis on finger.</content></item>
247 </list>
248 </section>
249 <section>
250 <caption>
251     <caption_cd V="1234-X" S="LOINC"/>Plan
252 </caption>
253 <list>
254     <item><content>Complete PFTs with lung volumes.</content></item>
255     <item><content>Chem-7</content></item>
256     <item>
257         <content>
258             Provide educational material on inhaler usage and peak flow self-monitoring.
259         </content>
260     </item>
261     <item>
262         <content>Decrease prednisone to 20qOD alternating with 18qOD.</content>
263     </item>
264     <item><content>Hydrocortisone cream to finger BID.</content></item>
265     <item><content>RTC 1 week.</content></item>
266 </list>
267 </section>
268 </body>
269 </levelone>

```