CDA for Common Document Types: Objectives, Status, and Relationship to Computer-assisted Coding

by Liora Alschuler

Abstract

This paper describes the clinical document architecture (CDA) and the CDA4CDT (clinical document architecture for common document types) initiative and suggests how adoption of these standards supports computer-assisted coding (CAC).

How CDA4CDT Supports CAC

CAC uses either structured input or natural language processing (NLP).[1] Today, the majority of critical patient information is unstructured in narrative notes, so CAC through NLP holds great promise, and it is in this area where the clinical document architecture (CDA) and the project known as CDA4CDT (clinical document architecture for common document types) have a role to play.[2]

Curiously, the aspects of language that are taught at the earliest age—how to recognize a sentence, a section, a paragraph—can be the most difficult from an NLP perspective. These familiar linguistic structures provide a context for information, and context is a critical factor in differentiating whether "discharge" is from a nose, and therefore a symptom, or from a hospital, in which case it signifies an end to a patient encounter. CDA has a rigorous context model that states that the CDA header defines the context for the complete document unless explicitly overridden within the body, section, or data element within the document.

In short, NLP needs to know what is data and what is metadata. The syntax of CDA is Extensible Markup Language (XML), which makes this division between data and metadata explicit.[3] NLP needs to know what is a section, so CDA helps because it defines the sections and subsections within a document. When NLP needs to know the specific subject of a section, CDA4CDT helps because it defines clinical documents down to the section and sometimes to the element level.

Clinical Document Architecture

Clinical Document Architecture, Release 2 (CDA R2) is a specification for the exchange of electronic clinical documents. CDA defines a clinical document as having the following characteristics:[4]

- "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 an organization entrusted with its care."
- "Potential for authentication: A clinical document is an assemblage of information that is intended to be legally authenticated."

- "Context: A clinical document establishes the default context for its contents."
- "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."[5]

A key use case for CDA is making these electronic documents available within an electronic medical record (EMR), in a personal health record (PHR), and within a broad-based exchange initiative such as a regional or statewide health information exchange network. CDA R2 is based on the Health Level Seven (HL7) Reference Information Model (RIM), an object model that creates a common framework for health information used in clinical care, research, public health, and reimbursement. CDA is at the core of successful health information exchange networks around the world.

CDA for Common Document Types

CDA4CDT is a privately-funded initiative of the healthcare transcription and document management industry. It was started by M*Modal, AHIMA, and the Association for Healthcare Documentation Integrity (AHDI; formerly the American Association for Medical Transcription and now affiliated with the Medical Transcription Industry Association [MTIA]). Since its initiation at the beginning of 2007, seven benefactors, one sponsor, and several provider participants have joined. (See the full list in the acknowledgments at the conclusion of this article.)

The mission of CDA4CDT is as follows:

- Develop CDA implementation guides for common types of electronic healthcare documents
- Bring CDA implementation guides through the HL7 ballot process
- Promote use and adoption of CDA implementation guides by healthcare organizations and health information exchange networks

Common document types are those typically produced through the transcription process: discharge summaries, history and physical (H&P) notes, consultation notes, and so on. A CDA implementation guide is a set of constraints on the base standard, CDA. So, for example, where the base standard allows any type of document section and sets no minimum requirements for report content, an implementation guide is specific to a report type and sets minimum requirements for the sections included in that type of report. The first report type defined by CDA4CDT is the History and Physical, which defines the sections shown in Table 1.[6]

In Table 1, "R" indicates that the section is required, while "O" indicates that the section is optional. Note that a separate list, Appendix D in the Draft Standard, includes sections for 51 typical physical examination subsections, including HEENT (head, eye, ear, etc.). The codes listed in the table are from the Logical Observation Identifiers, Names, and Codes (LOINC) terminology.[7]

The History of Present Illness (HPI) section might appear as shown in Figure 1 when displayed in a Web browser or printed.

Transcription systems can comply with the requirements of CDA4CDT using current word processing formats for the body of the clinical report. If a transcription system chooses to raise the value of the report, however, for integration with EMRs, it can output the report in XML. All current transcription systems use XML internally, and CDA is designed for automated, hands-off translation from the data already available to the transcription system. (See Figure 2.)

Note that the interface used by the transcriptionist need not show the XML markup. Transcription systems can convert word processing format to XML and can use native XML editors that provide a clean user interface.

As a simple example of the power of XML tagging, in the display format shown in Figure 2, the table of contents is automatically generated by looking for the following sequence and inserting the character string within the title element as an entry:

```
<section>...<title> ..... </title>
```

More complex applications of XML tagging permit the extraction of designated sections, for example, the entire HPI section, and its insertion into a discharge summary for review by the dictating physician. EMR applications will extract medication, problem, and allergy lists for integration into the electronic patient chart.

Incremental Semantic Interoperability

A key to understanding how CDA and CDA4CDT will affect the transcription industry and CAC is the concept of incremental semantic interoperability. This concept is predicated on the assumption that essential clinical information—information that is an integral part of the patient chart—will be produced by systems with heterogeneous levels of technical sophistication for some time to come. In other words, for the foreseeable future we will need to access, integrate, and communicate information that comes from highly coded EMRs, voice interface transcription, paper notes, and everything in between. Conversely, interoperability across systems will fail if it is predicated on an assumption of a uniform and high level of technical sophistication and coding.

There is little or no margin for error when communicating medical information in support of the delivery of care. Putting an emphasis on correct information requires that information exchange allow uncoded free text and narrative as well as handwritten notes. CDA is designed to support information exchange in this heterogeneous environment. A CDA document can therefore be as simple as a scanned note or word processed document attached to the following metadata fields:

- ID (globally unique identifier, typically inserted automatically by a software application)
- Document type (classifies the document as an encounter note, H&P note, consultation note, and so forth)
- Date/time of document publication (when available for patient care)
- Confidentiality code (typically “N” for normal; can be set higher or lower or use a local coding system)
- Patient (minimum is some kind of identifier such as a medical record number; can include the usual demographics such as name, date of birth, and so forth)
- Author (minimum is some kind of personal identifier and can also include name, demographics, and contact information)
- Custodian (the institution responsible for maintaining a copy of the document)

In a typical transcription workflow, all the required metadata would be imported automatically through the order interface, perhaps using a translation of an HL7 version 2.x message.

For simple file formats, CDA4CDT imposes minimum content requirements. (See Table 1 for the list of required sections proposed for H&P notes.) Where no information is available, the section would be inserted along with a string such as “No information given” that must be reviewed before signature. For systems capable of XML output, these sections would carry the markup defined in the specification. This markup provides much greater potential for reuse and for validating that minimal content requirements have been met. This approach to incremental interoperability also has ramifications for CAC.

Conclusion: Status of CDA4CDT

The CDA4CDT project has developed two implementation guides, one for the H&P note and one for the consultation note. As of this writing, these implementation guides are currently open for ballot as HL7

Draft Standards for Trial Use (DSTUs)—the ballot closes September 10, 2007. If they pass, they will be published within three months for trial implementation. If substantive changes are required because of the ballot review, the implementation guides will go back to ballot in January 2008. The next implementation guide to be developed will be the operative note.

The true challenge for CDA4CDT, however, lies beyond the formal ballot and publication process. The project will become useful in the real world and have an impact on CAC only when it is accepted by providers as a baseline requirement for transcription services and only when all applications that manage clinical documents—for CAC and for any other type of secondary usage—are compliant with the specification. The key benchmark for success of CDA4CDT, then, is when it becomes a common criterion for purchase and implementation of all electronic document software and services.

Acknowledgments

The CDA4CDT project was founded by M*Modal, the American Health Information Management Association (AHIMA), and the Association for Healthcare Documentation Integrity (AHDI; formerly the American Association for Medical Transcription (AAMT), now affiliated with the Medical Transcription Industry Association [MTIA]).

These founders have been joined by industry benefactors Spheris, MedQuist, InterFix, Precyse Solutions, Webmedx, MDinTouch, and 3M. Without their support and participation, this project would not have been possible.

In addition, this project has benefited from the participation of Acusis, Kaiser Permanente, Mayo Clinic, Military Health System, University of Pittsburgh Medical Center, and GE Healthcare. Project management was provided by Alschuler Associates, LLC.

Liora Alschuler is the founder of Alschuler Associates, LLC, in East Thetford, VT.

Notes

1. MedQuist IP Corporation. "The Impact of Computer-Assisted Coding, Focus: Coder Productivity and Roles." White paper. Mount Laurel, NJ, 2005. Available at http://medquist.com/pdf/coding/Coding%20Productivity%20Whitepaper.pdf.

2. Health Level Seven, Inc. "HL7 Clinical Document Architecture, Release 2.0, ANSI-approved HL7 Standard." Ann Arbor, MI, May 2005.

3. Dolin, Robert H., et al. "HL7 Clinical Document Architecture, Release 2." *Journal of the American Medical Informatics Association* 13, no. 1 (2006): 30.

4. http://www.regenstrief.org/medinformatics/loinc/

5. Health Level Seven, Inc. "HL7 Clinical Document Architecture, Release 2.0, ANSI-approved HL7 Standard." Section 1.1. Ann Arbor, MI, May 2005.

6. Health Level Seven, Inc. "HL7 Implementation Guide for CDA Release 2: History and Physical Notes (U.S. Realm)." Draft Standard for Trial Use, Release 1, Levels 1, 2, and 3. Ann Arbor, MI, August 2007.

7. http://www.regenstrief.org/medinformatics/loinc/

Table 1
CDA4CDT History and Physical Sections

Section Category	R/O	Code	Component Name
Reason for Visit/Chief Complaint	R	29299-5	REASON FOR VISIT
		10154-3	CHIEF COMPLAINT
		46239-0	REASON FOR VISIT + CHIEF COMPLAINT
History of Present Illness	R	10164-2	HISTORY OF PRESENT ILLNESS
Past Medical History	R	11348-0	HISTORY OF PAST ILLNESS
Medications	R	10160-0	HISTORY OF MEDICATION USE
Allergies	R	48765-2	ALLERGIES, ADVERSE REACTIONS, ALERTS
Social History	R	29762-2	SOCIAL HISTORY
Family History	R	10157-6	HISTORY OF FAMILY MEMBER DISEASES
Vital Signs	R	8716-3	VITAL SIGNS (may be a subsection of Physical Examination)
Review of Systems	R	10187-3	REVIEW OF SYSTEMS
Physical Examination	R	29545-1	PHYSICAL FINDINGS
		Optional Subsections	
		10210-3	GENERAL STATUS, PHYSICAL FINDINGS (optional, must be subsection)
		See Appendix D—List of Additional Physical Examination Subsections.	Additional optional subsections. List includes those in section 3.3.3 (Provider Unspecified History and Physical Note) of the "Additional Information Specification 0004: Clinical Reports Attachment."
Diagnostic Findings	R	30954-2	RELEVANT DIAGNOSTIC TESTS AND/OR LABORATORY DATA
Assessment and Plan	R	AAPLN-X	ASSESSMENT AND PLAN
		ASSMT-X	ASSESSMENT
		18776-5	PLAN
Procedure History	O	10167-5	PROCEDURE HISTORY
Immunizations	O	11369-6	HISTORY OF IMMUNIZATIONS
Problems	O	11450-4	PROBLEM LIST

Figure 1
Display of a CDA4CDT History and Physical Document

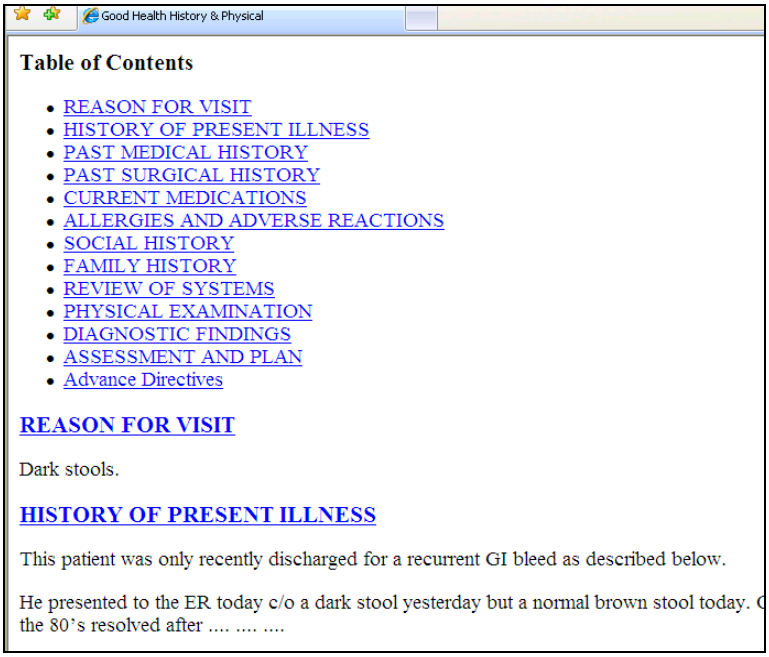

Figure 2
CDA4CDT XML for an HPI Section

```
<section>
    <templateId root="1.3.6.1.4.1.19376.1.5.3.1.3.4"/>
    <code codeSystem="2.16.840.1.113883.6.1"
       codeSystemName="LOINC" code="10164-2"
       displayName="HISTORY OF PRESENT ILLNESS"/>
    <title>HISTORY OF PRESENT ILLNESS</title>
    <text>
       <paragraph>This patient was only recently discharged for
a recurrent GI bleed as described below.</paragraph>...
```