

Case Report ■

Iterative Evaluation of the Health Level 7—Logical Observation Identifiers Names and Codes Clinical Document Ontology for Representing Clinical Document Names: A Case Report

SOOKYUNG HYUN, RN, DNSc, JASON S. SHAPIRO, MD, MA, GENEVIEVE MELTON, MD, MA, CARA SCHLEGEL, RN, MS, PETER D. STETSON, MD, MA, STEPHEN B. JOHNSON, PhD, SUZANNE BAKKEN, RN, DNSc

Abstract The authors summarize their experience in iteratively testing the adequacy of three versions of the Health Level Seven (HL7) Logical Observation Identifiers Names and Codes (LOINC) Clinical Document Ontology (CDO) to represent document names at Columbia University Medical Center. The percentage of documents fully represented increased from 23.4% (Version 1) to 98.5% (Version 3). The proportion of unique representations increased from 7.9% (Analysis 1) to 39.4% (Analysis 4); the proportion reflects the level of specificity in the document names as well as the completeness and level of granularity of the CDO. The authors shared the findings of each analysis with the Clinical LOINC committee and participated in the decision-making regarding changes to the CDO on the basis of those analyses and those conducted by the Department of Veterans Affairs. The authors encourage other institutions to actively engage in testing healthcare standards and participating in standards development activities to increase the likelihood that the evolving standards will meet institutional needs.

■ *J Am Med Inform Assoc.* 2009;16:395–399. DOI 10.1197/jamia.M2821.

Introduction

The HL7-Logical Observation Identifiers Names and Codes (LOINC) Clinical Document Ontology (CDO) and related Document Type Vocabulary Domain^{1,2} were developed to support the exchange of clinical documents. A CDO Task Force developed a white paper that specified the requirements for and initial draft of a polyhierarchical classification of documents that would support exchange of clinical documents. The CDO was designed to support the following use cases: (1) document location—retrieve pertinent information through a query for documents of a specified class; (2) document organization—present a logical manner of sorting or organizing documents for a variety of purposes; (3) document structure—enable categorization of information according to underlying

structure; (4) document display—predict display requirements such that documents classified alike share display characteristics. The Clinical LOINC Committee has responsibility for evolution and maintenance of the CDO.

As part of the process of developing an architecture that supports entry and analysis of clinical narratives (i.e., free or semi-structured texts created within the context of clinical care) for multiple purposes, we recognized the potential utility of a CDO to assist in creating and retrieving clinical documents and document templates at our institution.³ Consequently, we undertook an iterative process in which we tested the CDO using document collections from the Columbia University Medical Center campus (CUMC) of New York Presbyterian Hospital (NYPH) and contributed to the evolution of the CDO through sharing our findings with the Clinical LOINC Committee and participating in the decision-making process regarding CDO axis definitions and value sets. In this article, we report our testing of the adequacy of three versions of the CDO for representing document names at NYPH-CUMC and discuss the implications for refining the CDO and for improving document and document template naming conventions at our institution. Our case report not only illustrates the evolution of the CDO over time, but also serves as a general example of how an institution can work closely with a standards development group to inform the evolution of a standard that meets the needs of institutional use cases.

Case Description

The CDO is composed of four axes: Subject Matter Domain (SMD) (e.g., cardiology), Professional Level (e.g., registered nurse), Setting (e.g., Inpatient), and Type of Service (e.g., Subsequent evaluation). The fifth axis, Kind of Document, is

Affiliations of the authors: School of Nursing (SH, CS, CSB), Department of Biomedical Informatics (JSS, GM, PDS, SBJ, SB), Department of Medicine (PDS), Columbia University, New York, NY; Department of Emergency Medicine, Mount Sinai School of Medicine (JSS), New York, NY.

Supported by grants from the National Institutes for Health: 1R01 LM 07268 (S. Johnson, S. Bakken, S. Hyun), T15 NLM 007079 (J. Shapiro and G. Melton), 5K22 LM 008805 (P. Stetson) and T32 NR 007799 (C. Schlegel). The authors thank the domain experts for their participation and the other members of the eNote working group for their advice and support, as well as the HL7 DOTF and Clinical LOINC Committee for providing the foundation for the analyses in this Case Report and for their continuing work in this important area.

Correspondence: Sookyung Hyun, RN, DNSc, Columbia University School of Nursing, 630 W. 168th Street, Mail Box 6; New York, NY, 10032; e-mail: <sh2138@columbia.edu>.

Received for review: 04/08/08; accepted for publication: 01/14/09.

under development and its value list is limited to clinical documents.

We performed four analyses using three versions of the CDO: Original CDO (Version 1), Original CDO plus expanded values for SMD (Version 2), and CDO with expanded value list in all axes (Version 3). First, we tested CDO Version 1 using the data set of document names (notes and reports; $n = 894$) from our institutional data dictionary, the Medical Entities Dictionary (MED).⁴ Second, we expanded the SMD value list by integrating with the American Board of Medical Specialties' (ABMS)⁵ specialty names (Version 2) and compared the representation of note names ($n = 163$) using the initial SMD values and the expanded SMD values.⁶ Third, we examined the adequacy of CDO Version 2 for representing nursing document names ($n = 94$) from the Eclipsys XA documentation system and the MED. Last, we examined the adequacy of CDO Version 3 using an institutional data set ($n = 935$) inclusive of a non-redundant set of document names from Analyses 1–3. One author (Suzanne Bakken), a member of Clinical LOINC Committee, presented the results of each analysis in person or via teleconference to the Committee and participated in decision-making regarding changes to the CDO on the basis of these analyses and those conducted by the Department of Veterans Affairs. In Fall 2007, the Clinical LOINC Committee approved an expanded value list for each axis of the CDO (See <http://www.regenstrief.org/medinformatics/loinc/discussion-documents>). This list will continue to be refined through contributions from the user community.

Methods

Methods were similar across the four analyses. Research team members were trained before individually defining the document names and a team member (Suzanne Bakken) familiar with the development of the CDO reviewed the

initial coding and provided feedback to others. We used a set of coding guidelines created by the research team. If a document name did not include information related to an axis of the CDO, it was categorized as *Not Specified*. If a document name included information related to an axis of the CDO, but there was no appropriate value in the value list for the axis, it was categorized as *Other*.

Findings were summarized using percentages. Inter-rater agreement between coders was calculated using the kappa statistic on a random sample of documents. The percent of documents coded as *Not Specified* for a particular axis is a reflection of the granularity of the document names, not the completeness of the CDO. The percent of documents coded as *Other* for a specific axis is a measure of the completeness of the CDO value list for the axis. In regards to categorizing the document name in its entirety rather than by axis, the minimum requirement for a fully-specified document name at the time of these analyses was an appropriate value in Kind of Document axis (i.e., clinical document) plus one additional axis, thus a value of *Not Specified* for as many as three axes did not preclude designation of a document name as fully-specified. Document names that included a value of *Other* in at least one axis were coded as not fully-specified because the CDO did not contain an appropriate axis value for the document. The proportion of document names in an analysis that did not share one or more axis values with another document is reported as the percent of uniquely specified documents and is influenced by both the availability of appropriate CDO values and the granularity of the document names in the data sets.

Example

Results of the four analyses are summarized in Table 1. The proportions of fully-specified (23.4 versus 98.5%) and uniquely specified (7.9 versus 39.1%) document names in-

Table 1 ■ Summary of Analyses

CDO Version (year)	Analysis 1				Analysis 2		
	Version 1 (2000)				Version 2 (2005)		
Number of documents	$n = 894$ notes and reports				$n = 163$ notes		
Fully-specified documents	23.4%				—		
Uniquely specified documents	7.9%				—		
	% Specified with CDO Axis Value	% Specified with Other as Axis Value	% Not Specified	IRR Kappa (prob > Z)	% Specified with CDO Axis Value	% Specified with Other as Axis Value	% Not Specified
Subject matter domain	26.7	73.3	10.7	0.65 (0.0000) ^a ; 82.5% agreement ^{b†}	100	0	55.8
Professional level	99.9	0.1	90.8	0.97 (0.0000) ^a ; 100% agreement ^{b†}	—	—	—
Setting	99.9	0.1	90.2	0.97 (0.0000) ^a ; 100% agreement ^{b†}	—	—	—
Type of service	43.5	56.5	3.9	0.82 (0.0000) ^a ; 0.72 (0.0000) ^b	—	—	—
Kind of document	100	0	0.1	0.75 (0.0000) ^a ; 100% agreement ^{b†}	—	—	—

CDO = clinical document ontology.

Note [†]Unable to calculate kappa; ^aBetween Sookyoung Hyun and Cara Schlegel; ^bBetween Suzanne Bakken and Genevieve Melton; ^cbetween Suzanne Bakken and Sookyoung Hyun; Value list for Kind of Document is under development.⁹

creased from CDO Version 1 and Version 3. The proportions of documents that had a specified CDO axis value as compared to a value of *Other* also increased for SMD (26.7 versus 98.6%) and Type of Service (43.5 versus 99.9%) for CDO Version 1 versus Version 3. Values for Professional Level, Setting, and Kind of Document were sufficiently well-specified for our document names in CDO Version 1. The testing of CDO Version 2, which was augmented by SMD values from the ABMS, with clinical notes (Analysis 2) versus nursing documents (Analysis 3) showed a difference (100 versus 64.2%) in the percentage of documents with specified SMD values. Differences in % *Not Specified* among the four analyses are a reflection of the different document data sets used not of the changes in the CDO.

We calculated inter-rater reliability of the coding for two analyses. In Analysis 1, 208 (23%) document names were randomly selected. Kappa statistics ranged from 0.65 to 0.97 between Sookyung Hyun and Cara Schlegel. The most frequently occurring disagreement (18%) for SMD was related to Pathology versus *Other*. Agreement between Suzanne Bakken and Genevieve Melton on a small sample of document names was 100% in 3 of 5 axes. For Analysis 4, the kappa statistics for inter-rater reliability ranged from 0.57 (Type of Service) to 0.96 (Professional Level between Suzanne Bakken and Sookyung Hyun). For Type of Service, the most frequently occurring disagreements were related to Diagnostic Study versus Procedure (15%) and Operative procedure versus Procedure (3%). Table 2 provides the size of the value list for each axis and examples of the evolution of representation of NYPH-CUMC document names over the three versions of the CDO.

Discussion

The improvements in proportion of documents fully-specified and with unique representations from CDO

Version 1 to CDO Version 3 suggest that CDO Version 3 is superior in representing the NYPH-CUMC clinical document data set. The implication of the relatively large proportion of document names with non-unique representations even in Version 3 is that multiple documents will map to the same LOINC code. Some of this redundancy can be removed through more specific document names. In other instances, the value lists of SMD and Type of Service would require expansion; this is particularly true for nursing document names. The former is an institutional issue. Other organizations such as the Department of Veterans Affairs have undertaken extensive efforts in standardizing document names along with their efforts within the Clinical LOINC Committee to improve the CDO.⁷ The issue of expansion of value lists and how much of the expansion should be at the Clinical LOINC level versus the institutional level requires further discussion and testing of the CDO with document data sets from other institutions.

Another granularity-related issue identified in the analyses, particularly those related to nursing document names, is partially an artifact of the Eclipsys XA CIS documents in our data set. In some cases, document names reflected a fairly limited scope of contents that might be considered sections of other documents (e.g., "Eclipsys Note Event—MD Assessment Note: Social", "Nursing NICU Heart Sounds"). In addition, some types of nursing documents in our sample are represented as panels rather than documents in the LOINC database (e.g., "Nursing Progress Note: Braden Scale" and "Nursing Progress Note: Oral Assessment"). The Clinical LOINC Committee is the process of developing an editorial policy to handle such issues.

Table 1 ■ (continued)

Analysis 3			Analysis 4			IRR Kappa (prob > Z)
Version 2 (2005) n = 94 nursing notes 74.5%			Version 3 (2007) n = 935 notes and reports 98.5%			
33%			39.1%			
% Specified with CDO Axis Value	% Specified with Other as Axis Value	% Not Specified	% Specified with CDO Axis Value	% Specified with Other as Axis Value	% Not Specified	
74	26	29	98.6	1.4	5.8	0.85 (0.0000) ^c
100	0	0	100	0	85.7	0.96 (0.0000) ^c
100	0	0	100	0	86.4	0.89 (0.0000) ^c
100	0	4	99.9	0.1	18.2	0.57 (0.0000) ^c
100	0	0	99.1	0.9	0.1	0.96 (0.0000) ^c

Table 2 ■ Number of Values of the CDO Components and Sample Representations According to Version

	Version 1	Version 2	Version 3
CDO Component			
Subject matter domain	30	51	141
Professional level	23	19	29
Setting	12	12	18
Type of service	27	40	124
Kind of document	10	10	29
Example 1: Eclipsys Note Event— MD progress note: obstetrics-gynecology			
Subject matter domain	other	obstetrics and gynecology	obstetrics and gynecology
Professional level	physician	physician	physician
Setting	inpatient	inpatient	inpatient hospital
Type of service	evaluation and management	evaluation and management	progress note
Kind of document	clinical note explicit	document	clinical document
Example 2: CPMC Neurologic embolization*			
Subject matter domain	other	—	neurology
Professional level	not specified	—	not specified
Setting	not specified	—	not specified
Type of service	interventional procedure	—	operative procedure
Kind of document	report implicit	—	report
Example 3: Ideatel initial visit note			
Subject matter domain	other	diabetology	diabetology
Professional level	not specified	not specified	not specified
Setting	not specified	not specified	telehealth
Type of service	initial evaluation	initial evaluation	admission evaluation
Kind of document	clinical note explicit	document	clinical document
Example 4: Eclipsys Note Event— Nursing progress note: nutrition/metabolic			
Subject matter domain	other	other	nutrition/dietetics
Professional level	nurse	nurse	registered nurse
Setting	inpatient	inpatient	inpatient hospital
Type of service	evaluation and management	evaluation and management	progress note
Kind of document	clinical note explicit	clinical note	clinical document
Example 5: Eclipsys Note Event— Nursing progress note: Glasgow Coma Scale			
Subject matter domain	other	other	neurology
Professional level	nurse	nurse	registered nurse
Setting	inpatient	inpatient	inpatient hospital
Type of service	evaluation and management	subsequent evaluation	risk assessment and screening
Kind of document	clinical note explicit	clinical note	clinical document

CDO = clinical document ontology; CPMC = California Pacific Medical Center.

Note *A report. It was used for only Analysis 1 (Version 1) and 4 (Version 3).

Document names in EHR systems may influence the efficiency of clinician's retrieval of documents for review and document templates for creation of new documents⁸ and clinical document exchange and sharing. The CDO requirement for a fully-specified document name is limited to Kind of Document and at least one axis. For more informative and parsimonious document naming, it may be useful for the Clinical LOINC Committee to consider additional requirements for fully-specified document names.

Our work is the first formal comparison of the three versions of the CDO; however, there are several limitations to our analysis. The sample documents were from only one institution. Other settings and institutions may have different clinical document names. While comprehensive for our institution, the sample of document names

was relatively small. Moreover, as noted in our Case Description, our analyses of CDO Versions 1 and 2 were shared with the Clinical LOINC Committee, increasing the likelihood that values suitable for our documents would be included in CDO Version 3.

CDO Version 3 requires further testing to examine its suitability across multiple institutions and in multiple healthcare settings and to provide input for its expansion over time. The vision of a National Health Information Network that allows patient records to be securely, quickly, and easily transferred among clinicians and institutions across the country will require such an ontology. We encourage other institutions to actively engage in testing the CDO and other healthcare standards and to interact with standards groups to increase the likelihood that the evolving standards will meet institutional needs.

References ■

1. Document Ontology Task Force. Proposal for an Ontology for Exchange of Clinical Documents. Available at: <http://www.hl7.org/special/dotf/dotf.htm>. Accessed: Jan 18, 2008.
2. Document Ontology Task Force and the LOINC committee. HL7-LOINC Document Type Vocabulary Domain Paper. Available at: <http://www.regenstrief.org/medinformatics/loinc/discussion-documents/ontology.doc/view?searchterm=Document%20Type%20Vocabulary%20Domain>. Accessed: Jan 18, 2008.
3. Johnson S, Bakken S, Dine D et al. An electronic health Record based on structured narrative. *J Am Med Inform Assoc* 2008;15(1):54–64.
4. MED. Medical Entities Dictionary, 2005. Available at: <http://onto.cpmc.columbia.edu/medsite/med1.htm>. Accessed: Aug 18, 2005.
5. ABMS; ABMS. List of Approved ABMS Specialty Boards and Certificate, Evanston. Available at: <http://www.abms.org/>. Accessed: Feb 28, 2005.
6. Shapiro J, Bakken S, Hyun S, et al. Document ontology: Supporting narrative documents in electronic health records. *AMIA Annu Symp Proc* 2005.
7. Stevenson L. Document Requests from VA, Salt Lake City. *Clinical*, February 1 2008 LOINC Meeting.
8. Payne T, Kalus R, Zehner J. Evolution and use of a note classification scheme in an electronic medical record. *AMIA Annu Symp Proc* 2005.
9. Regenstrief. Institute. LOINC User's Guides. July 1 2008. Available at: <http://loinc.org/documentation>. Accessed: Jan 5, 2009.