

Standardizing Clinical Document Names Using the HL7/LOINC Document Ontology and LOINC Codes

Elizabeth S. Chen, PhD^{1,2}, Genevieve B. Melton, MD, MA^{4,5},
Mark E. Engelstad, DDS, MD^{4,6}, Indra Neil Sarkar, PhD, MLIS^{1,3}

¹Center for Clinical & Translational Science, ²Department of Medicine, ³Department of Microbiology & Molecular Genetics, University of Vermont, Burlington, VT

⁴Institute for Health Informatics, ⁵Department of Surgery, ⁶School of Dentistry, University of Minnesota, Minneapolis, MN

The standardization of clinical document names is an essential first step towards the optimal use, management, and exchange of documents within and across institutions. The HL7/LOINC Document Ontology (DO) is an existing and evolving document standard developed to provide consistent naming of clinical documents and to guide the creation of LOINC codes for clinical notes. The goal of this study was to explore the feasibility and challenges of mapping local clinical document names from two institutions into the five axes of the DO and then to leverage this mapping for identifying specific LOINC codes. The results indicate that the DO is either adequate or too broad for representing a majority of the document names and that there are LOINC codes available for one- to two-thirds of the names. Through this mapping process, granularity and other issues were revealed that will be valuable for guiding next steps towards effective standardization of clinical document names.

INTRODUCTION

Clinical documentation in the Electronic Health Record (EHR) plays a number of important roles ranging from facilitating diagnosis and treatment to enabling communication between caregivers. With the growing adoption of EHR systems, there is an increasing need for methods that improve access to the wealth of information from electronic documentation, including clinical notes. Establishing systematic and generalizable processes will be essential for incorporating standards that improve the search, retrieval, and use of clinical documents for patient care, quality initiatives, and research.

In the past decade, several efforts have emerged to facilitate the standardization, management, and exchange of clinical documents¹⁻⁶. The HL7 Clinical Document Architecture (CDA) is a standard for specifying the structure and semantics of a clinical document (e.g., clinical note or report)². This specification describes incremental levels for semantic interoperability from structuring the document header (metadata) to sections within the document to entries or statements within these sections. Among the required metadata elements for

the header is a “document type code” for specifying the kind of document where LOINC (Logical Observation Identifiers Names and Codes) has been specified as a source of these codes. The LOINC database provides a universal set of codes and names for laboratory and other clinical observations where each name can be represented according to six major axes: *Component, Property, Time, System, Scale, and Method*^{7,8}.

To address the use of heterogeneous document names by different EHR systems and/or institutions, the Document Ontology Task Force was charged with creating a system for consistent naming of documents to support a range of use cases (e.g., retrieval, organization, display, and exchange)³⁻⁵. The resulting HL7/LOINC Document Ontology (DO) consists of five axes for: *Type of Service, Kind of Document, Setting, Role, and Subject Matter Domain* where a fully specified name is considered one that includes a value from the Kind of Document axis and at least one of the other axes. Within LOINC, the five axes of the DO can be captured in three of the major axes (i.e., Component, System, and Method) to create a set of universal codes and names for clinical notes.

Several studies have been performed to evaluate the coverage of the DO⁹⁻¹¹ and LOINC codes¹² for representing local document names. These studies have led to enhancements to specific axes (e.g., Subject Matter Domain⁹) and have also described the need for extensions and clarifications for particular types of documents (e.g., nursing documents¹¹).

Previous work has focused on either coding local document names according to the axes in the DO or identifying LOINC codes for document names from a single EHR system or institution. The goal of this study is to build upon these efforts by exploring both the DO and LOINC codes for representing local document names from different EHR systems at two institutions. As part of a larger initiative to manage and exchange clinical documents across institutions, this study represents an initial step for gaining insights to the overall process, issues, and potential limitations associated with clinical document name standardization.

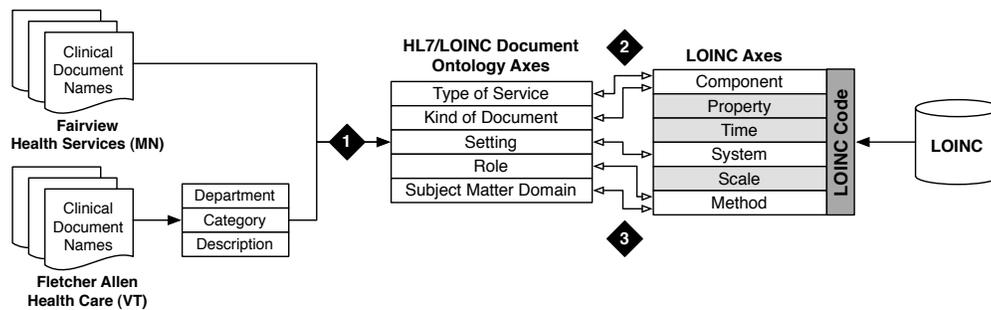


Figure 1: Overview of Methods.

METHODS

This study involved the collection, analysis, and mapping of clinical document names from two institutions: Fairview Health Services (FHS) and Fletcher Allen Health Care (FAHC). For sets of document names from both institutions, efforts were made to standardize the local names through a three-step process: (1) map each local name into the five axes of the DO, (2) map each LOINC code for clinical notes to the DO, and (3) map each local name to LOINC codes based on results from the first two steps (Figure 1). The overall goal of this study was to determine the level of coverage and challenges of using the DO and LOINC to represent document names at FHS and FAHC (with an initial focus on clinical notes) and provide insight into document standardization and exchange across institutions.

Collecting Clinical Document Names

FHS is a regional integrated health care network affiliated with the University of Minnesota¹³. The inpatient EHR system currently in use at its four metropolitan hospitals is Eclipsys, which has a range of functionality including electronic clinical documentation. All document names from the last five years were obtained from Eclipsys (Table 1A).

Table 1: Example Local Document Names.

A. FHS Document Names		
Cardiology Consult Service Daily Note		
Hospital Admission (H&P)-Attending		
Medicine Daily Note-progress note		
Nursing Progress Note-Admission from ER		
Surgery Daily Note-Post-Op Note		
B. FAHC Document Names		
Department	Category	Description
ANESTHESIOLOGY	Eval	Eval, H&P IP
CARDIOLOGY	Progress Note	Prog Note, OP
Gastroenterology	Consult	Consult, ED
General Surgery	Eval	Eval, Preop H&P
MEDICINE	Summary	Summary, Discharge

FAHC, affiliated with the University of Vermont, is the only tertiary care academic medical center in the State of Vermont¹⁴. Among the legacy electronic applications is a core clinical information system

containing inpatient and outpatient data since 1999 (e.g., transcribed notes). For this study, the names used for legacy transcribed notes from June 2008 to June 2009 were obtained where each name consists of three distinct parts for Department, Category, and Description (Table 1B).

Step 1: Mapping Local Names to the DO

As a preparatory step for mapping, all values in each DO axis were extracted and assigned “tree” codes that seek to preserve the hierarchies (up to 4 levels)⁵. At the time of this study, the DO values listed in the LOINC Users’ Guide⁷ and on the LOINC Web site⁵ were equivalent, and codes from the latter source were used. For example, tree code “9.k.l.” represents “Progress Note”, which is in the third level of the hierarchy in the Type of Service axis under “9.k.” for “Subsequent Evaluation” (second level) and “9.” for “Evaluation and Management” (top level)⁷.

Clinical document names in the FHS and FAHC sets were reviewed and mapped according to each axis in the DO. Mapping for each axis involved specifying a rating reflecting coverage within the DO (*adequate, too broad, too specific, not covered, or unspecified* [in the document name]) and most granular tree code (if possible). In addition, suggestions and other comments were provided, particularly in cases where a rating of adequate or unspecified could not be given or multiple codes from a single axis could be assigned. For the FHS set, rating and coding was performed directly for each document name; for the FAHC set, rating and coding was performed separately for unique values in each name part (Department, Category, and Description), which were then reassembled into the full document name. Interrater reliability was calculated based on mapping a subset of 50 names by two reviewers for each set.

Step 2: Mapping LOINC Codes to the DO

The LOINC 2.30 database was downloaded and processed to identify all codes designed according to the DO model (i.e., where CLASS=DOC.CLINRPT). For the LOINC codes identified, values in the Component, System, and Method axes were extracted

(the values for the other three major axes are the same across the codes: Property=Find, Time=Pt, and Scale=Doc). Similar to the coding of FAHC document names, unique values for the Component, System, and Method axes were mapped to values in the corresponding five axes in the DO, assigned the corresponding tree code, and then reassembled for each LOINC code.

Step 3: Coding Local Names with LOINC via DO

With the FHS and FAHC document names mapped to the DO (Step 1) and LOINC codes mapped to the DO (Step 2), the last step involved coding each local document name with a LOINC code by leveraging the DO mappings. First, basic automated methods were developed and used to identify codes through exact matching where values in each axis should be the same including null values (exceptions included unspecified values, which could match to any value). If an exact match could not be identified, an attempt was made to find a code based on matching of parent codes within the respective hierarchy in the DO.

RESULTS

The sets of document names from FHS and FAHC included 566 and 674 unique names respectively. For FAHC, these names consisted of 75 unique values for *Department*, 6 for *Category*, and 56 for *Description*. As of March 2010, the DO included 119 values for *Type of Service*, 19 for *Kind of Document*, 18 for *Setting*, 27 for *Role*, and 149 for *Subject Matter Domain*. Within LOINC, 327 codes were identified for clinical notes with 97 unique values for *Component*, 13 for *System*, and 96 for *Method*.

The distribution of ratings and inter-rater reliability from mapping FHS and FAHC document names to the DO can be found in Table 2A-B. In addition, the percentage of values from each DO axis covered in the local names can be found in Table 2C. Overall, the DO axes were either found to be adequate or too broad for the majority of cases (where the information was specified in the name). For both institutions, *Kind of Document* could be specified for almost all the names (given the focus on clinical notes, most mapped to “Note”). With respect to *Role*, most document names from both institutions did not include this information (in cases where this information was explicitly stated, the *Role* axis was found to be adequate). *Setting* was also found to be adequate for a majority of the names where this information was specified or could be inferred. In a small percentage of cases, the values were either too broad (e.g., the closest match for “Counseling Ctr Initial Assessment-FUMC” was “Outpatient”) or not included (e.g., “Hospice Note”).

The *Type of Service* and *Subject Matter Domain* axes involved the most cases where information within the document name was more specific than an existing axis value (“granularity mismatches”) or not covered. For example, “Interim Summary” maps to a less specific *Type of Service* value of “Summarization” and “Inpatient Oncology Surgery Consult-Initial” maps to the less granular *Subject Matter Domain* value of “Surgery”.

The coverage of values from each DO axis in the relevant LOINC codes is reported in Table 2C. In mapping the names associated with these LOINC codes to each DO axis, some mapping issues were encountered as exemplified in Table 3 (e.g., in addition to exact one-to-one matches, other types of matches were found including one-to-many and many-to-one). Resolving these issues was out-of-scope for this study; however, it is worth noting that this could have implications on accurate coding and harmonization efforts are on-going by the Clinical LOINC Committee⁵.

Table 2: Results from Mapping FHS, FAHC, and LOINC Names to Each DO Axis

	TOS	KOD	Setting	Role	SMD
A. Rating (FHS=no shading; FAHC=grey shading)					
Adequate	54%	99%	68%	4%	50%
	93%	100%	68%	-	92%
Too Broad	33%	-	2%	1%	21%
	3%	-	<1%	-	6%
Too Specific	1%	-	-	-	-
	1%	-	5%	-	-
Not Covered	2%	<1%	<1%	<1%	7%
	-	-	-	-	<1%
Not Specified	10%	<1%	29%	95%	21%
	3%	-	27%	100%	3%
B. Inter-Rater Reliability					
Proportion Agreement	92%	100%	90%	90%	100%
	100%	100%	94%	N/A	98%
Cohen's Kappa	0.95	1.00	0.84	1.00	0.82
	1.00	1.00	0.88	N/A	0.91
C. Coverage of DO Values					
FHS	28%	21%	28%	30%	42%
FAHC	10%	11%	33%	-	35%
LOINC	71%	11%	50%*	22%*	34%

TOS=Type of Service, KOD=Kind of Document, SMD=Subject Matter Domain, *=excludes unspecified Setting and Role values

Table 3: Examples of Mapping LOINC to the DO

Type	LOINC Name	DO Mapping
Exact 1-1	Component=Consultation Note	TOS=Consultation KOD=Note
Inexact 1-1	Method=Cardiology	SMD=Cardiovascular Disease
1-to-Many	Component=Education Procedure Note	TOS=Education TOS=Procedure KOD=Note
Many-to-1	Method=Gerontology Method=Geriatric Medicine	SMD=Geriatric Medicine
One-to-All	System={Setting}	Setting=Any value
No Match	Method=Substance Abuse	

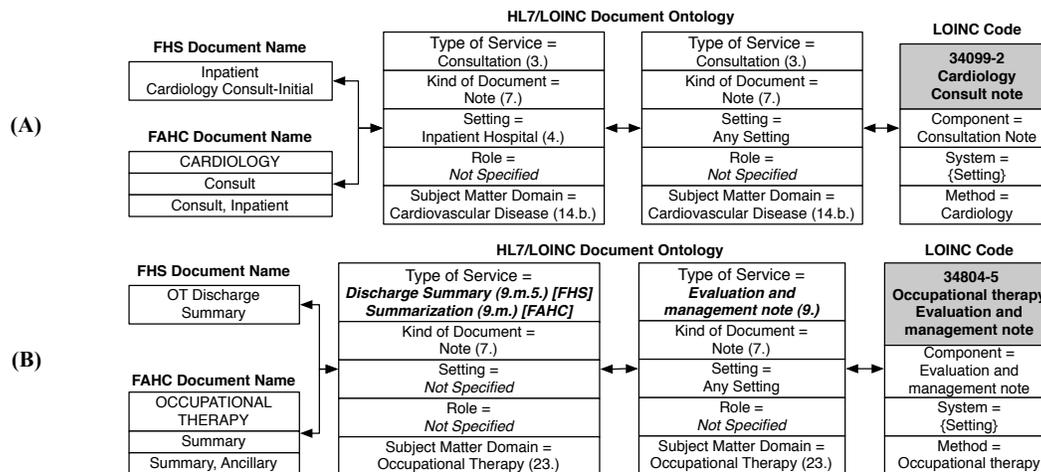


Figure 2. Mapping Local Document Names to the HL7/LOINC Document Ontology to LOINC Codes. (A) Exact match for FHS and FAHC document names; (B) Hierarchical match involving three levels in the Type of Service axis of the DO (bold italics) where the value for FHS is at the third level (tree code = 9.m.5.), FAHC value is at the second level (tree code = 9.m.), and LOINC maps to the top level (tree code = 9.).

Finally, in leveraging the DO mappings as part of the initial automated coding effort, LOINC codes for 25% of the FHS document names were identified based on exact matches, which increased to 63% when incorporating the DO hierarchies. For the FAHC set, LOINC codes could be assigned to 30% of the document names with exact matching, which increased to 47% with hierarchical matching. Figure 2 provides examples of exact and hierarchical mapping of local document names from both institutions to specific LOINC codes through the DO.

DISCUSSION

In this paper, we have described a process for mapping local document names to the five axes of the DO and then to specific LOINC codes. At each step, we have provided high-level statistics and observed potential challenges (e.g., granularity issues, inconsistencies and redundancies, and loss of information and meaning). Similar to findings from other studies that have examined the adequacy of the DO, we found that the existing values in each axis were able to represent a majority of the clinical document names. In addition, we were able to identify existing LOINC codes for one- to two-thirds of the names depending on the approach used.

Through the process used in this study, we have identified important mapping issues due to the nature of the local document names as well as the coverage of the DO and LOINC codes. We have also observed loss of information and meaning when mapping local document names to universal names and codes that could significantly impact effective access and use. For example, the inpatient document name from FHS

for “Inpatient Oncology Surgery Consult-Initial” maps to “Inpatient Hospital” for Setting and “Surgery” for Subject Matter Domain (thus losing information about oncology) and then maps to “General surgery Consult Note” in LOINC (further losing the explicit setting information). Depending on the use case, exact mappings may be essential (e.g., for seamless exchange between EHR systems) while broad mappings may be sufficient for other scenarios (e.g., retrieval of particular groups of documents). Additional work is thus needed to better understand the gaps and implications of applying these existing document standards. For example, there may be benefit in maintaining the original local document name and multi-axial DO representation in addition to the pre-coordinated LOINC code when exchanging documents in order to minimize information loss, preserve semantics, and enable flexible use.

Several types of granularity mismatches were revealed through preliminary analysis of cases where the Subject Matter Domain axis was found to be too broad. In addition to the “Oncology Surgery” example provided earlier, other similar examples include “Urogynecology Consult, OP” and “Cardiac Rehab Evaluation”. Additional examples include document names grouped into broader categories for this axis such as “Hematology and Oncology” (e.g., “Hematology Consult-Initial” and “Oncology Consult-Initial” from the FHS set). In addition, certain types of notes have two possible foci. For example, “Pediatric Speech Language Pathology Progress Note” loses significant meaning and context if mapped to either “Pediatrics” or “Speech Pathology” and also points towards additional

complexity that can occur with healthcare delivery. These examples illustrate issues resulting from pre-coordination of values and whether a post-coordinated approach would be valuable (e.g., adding “Urogynecology” in one or more hierarchies versus selecting existing values of “Urology” and “Gynecology”). Either approach could have implications on the retrieval of relevant documents.

While the Subject Matter Domain axis includes a value for “Multi-specialty Program” that could be used to represent “Breast Care Center” and “Sleep Center Evaluation”, this results in significant information loss. The Document Naming Nomenclature (DNN) represents another effort for structuring clinical document names to improve access to information⁶. In creating the DNN, seven categories were identified (*Multidisciplinary Care Unit, Service, Section, Role, Event, Event Qualifier, and Episode of Care*) for describing information within a name. The first three categories represent different types or levels of care providing units and may be an approach for differentiating specialties, sub-specialties, and multidisciplinary units.

In addition to further analyzing and understanding the results from this study, additional steps also include refining and advancing the techniques towards a more automated approach for document mapping¹⁵. The identification of common synonyms or lexical variants may be one step in facilitating this where each of the sources studied could be used to seed such lists. Other next steps include determining how to handle inconsistencies and redundancies (e.g., LOINC codes with the same DO axis values such as 34771-6 [*General surgery Consult note*] and 34847-4 [*Surgery Consult note*]), studying the existing hierarchies and polyhierarchies in the DO in more depth (e.g., “Pain Management” and “Geriatric Medicine” have multiple parents), and further exploring the coverage of LOINC (e.g., no codes for clinical notes related to Pediatrics are available) and other LOINC codes that may be applied to clinical documents (e.g., those with Scale=Doc).

The Clinical LOINC Committee has been the forum for collaborative harmonization, development, and extension to the DO for over five years. Primary effort has been focused on revising the DO axis values in order to provide a framework for creation of additional needed LOINC codes for document names. In addition to potential contributions to this universal effort, the insights from the present study may also be valuable for guiding naming conventions and practices locally in order to facilitate document management and exchange within and across institutions.

Acknowledgments

The authors thank Rhonda Kost and Charles Roger for their valuable assistance in obtaining the clinical document names for FAHC and FHS respectively.

References

1. Payne TH, Graham G. Managing the life cycle of electronic clinical documents. *J Am Med Inform Assoc.* 2006 Jul-Aug;13(4):438-45.
2. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc.* 2006 Jan-Feb;13(1):30-9.
3. Frazier P, Rossi-Mori A, Dolin RH, Alschuler L, Huff SM. The creation of an ontology of clinical document names. *Stud Health Technol Inform.* 2001;84(Pt 1):94-8.
4. Document Ontology Task Force and the LOINC Committee. HL7 LOINC Document Type Vocabulary Domain. Available at: <http://loinc.org/discussion-documents/ontology.doc>. Last accessed July 12, 2010.
5. LOINC. Document Ontology. Available at: <http://loinc.org/discussion-documents/document-ontology>. Last accessed July 12, 2010.
6. Brown SH, Lincoln M, Hardenbrook S, Petukhova ON, Rosenbloom ST, Carpenter P, et al. Derivation and evaluation of a document-naming nomenclature. *J Am Med Inform Assoc.* 2001 Jul-Aug;8(4):379-90.
7. McDonald C, Huff S, Mercer K, Hernandez J, Vreeman DJ. LOINC Users' Guide-December 2009. Available at: <http://loinc.org/downloads/files/LOINCManual.pdf>. Last accessed March 15, 2010.
8. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem.* 2003 Apr;49(4):624-33.
9. Shapiro JS, Bakken S, Hyun S, Melton GB, Schlegel C, Johnson SB. Document ontology: supporting narrative documents in electronic health records. *AMIA Annu Symp Proc.* 2005:684-8.
10. Hyun S, Shapiro JS, Melton G, Schlegel C, Stetson PD, Johnson SB, et al. Iterative evaluation of the Health Level 7--Logical Observation Identifiers Names and Codes Clinical Document Ontology for representing clinical document names: a case report. *J Am Med Inform Assoc.* 2009 May-Jun;16(3):395-9.
11. Hyun S, Ventura R, Johnson SB, Bakken S. Is the Health Level 7/LOINC document ontology adequate for representing nursing documents? *Stud Health Technol Inform.* 2006;122:527-31.
12. Dugas M, Thun S, Frankewitsch T, Heitmann KU. LOINC codes for hospital information systems documents: a case study. *J Am Med Inform Assoc.* 2009 May-Jun;16(3):400-3.
13. Fairview Health Services. Available at: <http://www.fairview.org/>. Last accessed July 12, 2010.
14. McDowell SW, Wahl R, Michelson J. Herding cats: the challenges of EMR vendor selection. *J Healthc Inf Manag.* 2003 Summer;17(3):63-71.
15. Vreeman DJ, McDonald CJ. Automated mapping of local radiology terms to LOINC. *AMIA Annu Symp Proc.* 2005:769-73.