The Creation of an Ontology of Clinical Document Names

Pavla Frazier[a], Angelo Rossi-Mori[b,f], Robert H. Dolin[c,f], Liora Alschuler[d,f], Stanley M. Huff[e,f]

[a]University of Utah, Salt Lake City, Utah, USA
[b]Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche,Roma, Italy
[c]Kaiser Permanente, Pasadena, California, USA
[d]alschuler.spinosa, Thetford, Vermont and La Jolla, California, USA
[e]Intermountain Health Care, Salt Lake City, Utah, USA
[f]Document Ontology Task Force, Health Level Seven, Ann Arbor, Michigan USA

Abstract

The efficient use of documents from heterogeneous computer systems is hampered by differences in document-naming practices across organizations. Using an open-consensus method, the Document Ontology Task Force[1], with support from the Veterans Health Administration, addressed this pervasive problem by developing a clinical document ontology. Based on the analysis of over 2000 clinical document names, the ontology was used to formulate a terminology model which is currently being used to guide the creation of fully-specified document names in LOINC (Logical Observations, Identifiers, Names & Codes). Incorporation into LOINC will enable homogenous management of documents in a widely distributed environment and will also give rise to a rich polyhierarchy of document names.

Keywords:

Document; Documentation; Names; Terminology; LOINC; Ontology

Introduction

What's in a name? that which we call a rose
By any other name would smell as sweet -Shakespeare

Today, millions of narrative clinical documents are being sent daily between various health care information systems. These documents all have names, selected by either the author or as dictated by institutional policy. However, person to person, system to system, and across enterprises, the names are not consistent. System A may call a given document a "Discharge Summary" and System B might call

the same kind of document a "Summary of Inpatient Stay." The increased exchange of information among different organizations produce collections of persistent narrative documents from multiple heterogeneous sources. Mapping local, arbitrarily assigned, free text document names to universal codes enables care providers and researchers to find, retrieve and process documents within those collections efficiently.The naming of documents is just one aspect of much deeper issues that have been discussed for centuries: the meaning of meaning [1] and the development of categories, classifications, nomenclatures and ontologies. If representation of meaning through a shared vocabulary is the "heart and soul" of a clinical information system [2], then semantic interoperability, or the ability to effectively communicate meaning, is vital to the fulfillment of the promise of informatics as we move into the new century.Recognizing the need for improved information access, the US Veterans Health Administration provided support for Health Level Seven (HL7) [3], a standards developing organization accredited by the American National Standards Institute (ANSI), to convene a task force charged with the expressed purpose of creating a system for document naming. The desired system would assist in the homogeneous management of narrative documents from heterogeneous computer systems. The group became known as the Document Ontology Task Force (DOTF) and its work has resulted in the design and preliminary testing of a terminological model for document names. This model is currently being used to guide the preparation of a polyhierarchy of document names for incorporation into LOINC, as a model for the synergistic development of health information exchange standards and coded medical vocabularies [4]. Furthermore, it has been proposed that the resultant polyhierarchy also be mapped into SNOMED RT (Systematized Nomenclature of Human and Veterinary Diseases Reference Terminology). Once LOINC codes are issued, they can be a common point of reference for widely distributed applications. The purpose of this paper is to describe the design and testing of the current document model. The description and analysis of the actual polyhierarchy of document names will be

---

[1] Document Ontology Task Force: Stanley M. Huff (chair), Liora Alschuler, Calvin Beebe, Paul Carpenter, Robert H. Dolin, Leo Fogarty, Angelo Rossi Mori, Peter Schloeffel, Kent Spackman, John Spinosa, Karen Van Hentenryck, Steve Wagner

elaborated upon in a future paper.

Several groups are working in areas closely related to document naming. The National Health Service [5] in the United Kingdom has approached the problem from a bottom-up method, focusing on the standardization of document headings, also known as sections. Such work is essential and complementary to the standardization of document names. LOINC and DOTF members are working together on standardizing section names as well. Also, an effort by several researchers with the US VAMC to create and test document naming guidelines for was described at the October 2000 Clinical LOINC[2]/DOTF meeting [6].

The European Standardization Body (CEN) has been making progress on the standardization of document categories. CENs work is conceptually quite close to that of the DOTF, and in fact, has been a source of information for the DOTF. Table 1 shows a partial view of the table CEN published in ENV 13606-2 [7] for "document categories." It is a small set of codes for very generic document types, to be used as additional, standard codes in conjunction with local names and codes. The table represents one method for managing collections of exchanged documents. In contrast, the DOTF focused on building a model to systematically populate a large set of fully-specified names.

Table 1 - Sample Document Categories in ENV 13606-2

UID	Generic Document	Description
DTC 01	Notes on Consultations	Usually a complete record of a patient encounter or healthcare activity
DTC 02	Progress Notes	Usually a brief follow-up review or statement of outcome that may be difficult to interpret without the preceding entries
DTC 03	Notes on Procedures	Usually a specialist summary of a procedure (whether invasive or not), operation etc.
DTC 04	Tests/ Procedure Requests	A request for a specialist service without a transfer of clinical care (which remains with the requestor)
DTC 12	Care Plans	These present outstanding or intended activities, possibly dynamically generated from certain process status attributes of the underlying entries

---

[2] Clinical LOINC committee: Suzanne Bakken, James Barthel, Dean Bidgood, Bruce Bray, James Campbell, Bill Francis, Alan Golichowski, Karl Hammermeister, Stan Huff, Doug Martin, Clem McDonald, Dan Pollock, Angelo Rossi-Mori, Jeff Suico, Anders Thurin, Wayne Tracy, Pat Wilson

Materials and Methods

Design Process

Representing a broad range of interests in document exchange and a deep level of understanding of terminology precedent and practices, the members of the DOTF exchanged source material and proposals via email and discussed preliminary statements of requirements and use cases for several months. At an initial meeting in Vancouver in June 2000, a consensus was reached in terms of defining conceptual axes that disambiguate document names for the purpose of document retrieval and use. The group adopted a strategy of using a multiaxial approach to create document names where each name is derived from a set of terminology axes based on the assumption that such a model is necessary for understanding and using documents that have been exchanged between systems.

DOTF members performed an empirical analysis of over 2000 clinical document names from Mayo Clinic Medical Center in Rochester, Minnesota, 3M/Intermountain Health Care and the Veterans Administration Medical Center (VAMC) in Salt Lake City, Utah and, later, from the VAMC in Nashville, Tennessee. Through an informal process, the frequency of items found by the task force members led to the creation of the categories of terms and subsequent axes.

The results were described in a white paper titled "Proposal for an Ontology for Exchange of Clinical Documents" [8] which was made available in July 2000 for public comment. It was agreed that the existing LOINC names and codes for radiology documents would be reviewed and subsequent recommendations would be made to the LOINC committee. In October 2000, the DOTF and the Clinical LOINC committee held a two-day joint meeting in order to discuss and revise the DOTF model and to formulate the method for incorporation of the model into the LOINC structure, thereby enabling the actual population of the polyhierarchy. This paper reflects the revisions and corrections from the October meeting.

Ontology Scope

The scope of work for the DOTF is defined by its use cases and requirements.

Use Cases

The ontology is intended for use in the following of situations:

- retrieval. The ontology should make possible retrieval of pertinent information through a query for documents of a specified class.

- organization. The ontology should provide a logical manner of sorting or organizing documents for a variety of purposes, including organizing and distributing the work of writing document schema definitions.

- preparation of templates. The ontology should make possible the preparation of templates for

categorization of information according to an underlying structure, although this may not itself be an underlying principle of classification.

- display. The ontology should be a good predictor of display requirements such that documents classified similarly share display characteristics, as a consequence of the internal structure.

Two types of requirements were applied: general vocabulary requirements and project-specific requirements.

General vocabulary design requirements

- Define all terms and concepts

- Compose document classes from primitive axes.

- Associate a "fully-specified name" and a "common usage name" to every class. The fully-specified name must be unique

- The ontology will allow creation of a polyhierarchy, which means that a given class may have multiple parents.

- Assign a unique, meaningless code to every document class.

- Where possible, use existing controlled vocabulary to define primitive axes.

Project-specific design requirements

The multiaxial naming strategy should be able to give unique names to clinical documents in use in clinical systems. The set of covered documents is defined empirically by lists supplied by participants and as defined in the HL7 Clinical Document Architecture (CDA) Specification, formerly known as the Patient Record Architecture (PRA). The CDA is a document markup ANSI-approved HL7 standard that specifies the structure and semantics of clinical documents for the purpose of exchange [9].

- The ontology must be the basis for systematic definitions of document classes to support the uniform management of clinical documents from heterogeneous sources.

- The name has a definite relationship to an activity (e.g. a visit) or an event (e.g. a medical error) being documented, but the document name may not be completely specified by the name of the activity or by the kind of event.

- The name may reflect the context (or some aspect of context) in which the document was created.

- The name may overlap and be redundant with information provided within the document header or within the document body.

- The fully-specified names have a bearing on the contents of the document. Document class does not define contents sufficiently for arbitrary machine processing, but is a sufficient predictor for human understanding.

While the fully-specified name does not, in and of itself, completely specify the contents of a document, it does set an expectation of what is to be found in the document and is based on expected document contents. Thus, while a "Colonoscopy report" is literally a report about a colonoscopy, there is no guarantee that information related to other procedures may or may not be found in the same document.

Domain development

Each axis draws from a designated domain. As the model was developed, it became apparent that some of the axes did not have pre-existing value sets. Therefore new values for several domains have been developed. Several domains are in the process of being incorporated into the HL7 reference information model (RIM) [10]. Ideally, each domain would provide all possible values for the associated axis. Moreover, the taxonomies within each axis will be the basis to produce the polyhierarchy of the fully-specified document names.

Results

Terminological Model

The proposed model presents a set of rules for use of six primitive axes that are represented in figure 1. Each axis corresponds to a specific domain, from which the descriptors will be taken, to generate the fully-specified document names.
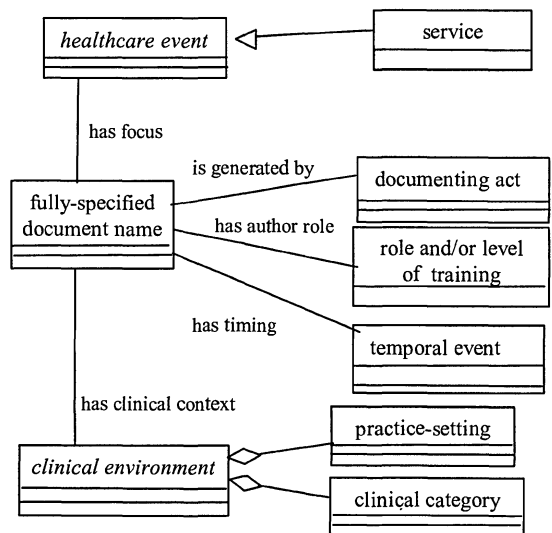

Figure 1- Graphical representation of the terminological model in Unified Modeling Language

Table 2 – Segments and sub-segments to be represented in the LOINC structure

Axis	Proposed domain	Examples of descriptors
Service	SNOMED RT Procedure axis [11], for example	consult, procedure, evaluation and management
Documenting Act	New	note, report, summary
Temporal Event	New	admission, discharge, birth, transfer
Clinical Category	Clinical category classification HL7 RIM (proposed)	cardiology, renal transplant, legal medicine
Practice Setting	HL7 RIM	hospital unit, out-patient clinic
Kind of Narrative	New	whole document, document section
Role/level of training of Documenter	New	attending physician, nurse practitioner, medical student

Rules for a Document Ontology -Primitive Axes

The service axis is the only required axis and may contribute more than one descriptor to form the fully-specified name. The remaining axes can be represented at most once. The rules are for specifying the meaning of a class and do not represent local or conventional names. However, which names will be displayed for the end-user, the fully-specified or common usage names associated with the defined classes, is assumed to be a local implementation decision.

- service. The basic high-level act being documented.
- documenting act. Type of documentation done.
- temporal event. Time aspect of topic of document.
- clinical category. The clinical domain of practice.
- practice setting. A categorization of the clinical setting in which care is delivered. This refers to the type of facility, not the physical location.
- kind of narrative. Refers to whether it is the entire document or part of one, i.e. section.
- role of documenter. Role is in relationship to the subject (patient, family) of the document and in terms of level of training (resident, student nurse, etc.)

Representation in the LOINC structure

The segments and sub-segments that are proposed to be represented in the LOINC structure are described in table 2. A few examples of fully-specified names corresponding to that structure are presented in table 3.

Discussion

Although the main focus for the development of the ontology was on dictated documents, the overall scope of application of the ontology should be wider. Given the pragmatic approach of LOINC, and its effective revision cycle, the final LOINC list will end up in accommodating all the persistent documents appearing in routine collections from heterogeneous sources that must be retrieved and organized in real settings. As we apply classification to additional classes of documents, we may discover the need for additional axes. Future work that makes use of resulting document polyhierarchy may chose to tie contents to class more strongly than is possible when applying a classification system to existing, heterogeneous documents. For example standard templates could define typical section and subsection titles for classes of homogeneous documents, in order to facilitate dictation and transcription, and to assure a complete and systematic generation and presentation of the documents. The ontology could then be used to name and organize large classes of documents that share the same template.

Table3 – Document mapping to DOTF Model

Original Document Name	Clinical Category	Temporal Event	Service	Documenting Act	Practice Setting	Role /level of Training of Documenter	Kind of Narrative
Transplant NP clinic notes	Transplanta-tion surgery		evaluation and management	note	transplant clinic	nurse practitioner	document
Cardiac cath report in-patient	Interventional cardiology		cardiac catheterization	report	hospital		document
CV surg fellow DC summary	Cardiovascular surgery	discharge	summation	summary		fellow	document
Derm consult	Dermatology		consult				Document
DM clinic education note	Diabetes		education. diabetes	note	clinic		Document

Acknowledgments

References

[1] Ogden CK and Richards IA. *The Meaning of Meaning.* New York: Harcourt, Brace & World, Inc., 1923.

[2] Huff SM, Craig RB, Gould, BL, Castagno DL and Smilan RE. Medical data dictionary for decision support applications. In: Stead WW, ed. *Proc 11ᵗʰ SCAMC.* Los Angeles: Computer Society Press, 1987; pp. 310-17.

[3] Health Level Seven. http://www.hl7.org

[4] Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD Jr, Forrey AW, Francis WG, Tracy WR, Leavelle D, Stalling F, Griffin B, Maloney P, Leland D, Charles L, Hutchins K, and Baenziger J. Development of the LOINC (Logical Observations Identifiers, Names, and Codes) vocabulary. *J Am Med Inform Assoc.* 1998: 5 (3) pp: 276 -292.

[5] Position paper: Towards an information standard for organizing clinical communications, National Health Service, http://www.hl7.org/special/dotf/docs/Position Paper: 2000: pp. 1-23.

[6] Brown SH and Lincoln M. VAMC Document Name Presentation, 10/24/00 DOTF/LOINC Meeting, Salt Lake City UT.

[7] A1ENV13606-2, Health Informatics Communication of Electronic Health Care Record part 2, Domain Termlist, CEN, 1999.

[8] Proposal for an ontology for exchange of clinical documents, Document Ontology Task Force, http://www.hl7.org/special/dotf/: 2000: pp.1-13.

[9] Alschuler L, Dolin RH, Boyer S and Beebe C, eds. Clinical Document Architecture Framework, Release 1.0. HL7, 2000.

[10] HL7 Reference Information Model. www.HL7.org: 2000.

[11] Systematized Nomenclature of Medicine. SNOMED RT. www.snomed.org.

Address for correspondence

Correspondence: Pavla Frazier, MSN MBA, Department of Medical Informatics, University of Utah, Salt Lake City, UT 84132; e-mail: lppfrazi@ihc.com.