Cross-mapping clinical notes between hospitals: An application of the LOINC Document Ontology

Li Li, MS, C. Paul Morrey, PhD, David Baorto, MD, PhD
New York Presbyterian Hospital, New York, NY

Abstract

Standardization of document titles is essential for management as the volume of electronic clinical notes increases. The two campuses of the New York Presbyterian Hospital have over 2,700 distinct document titles. The LOINC Document Ontology (DO) provides a standard for the naming of clinical documents in a multi-axis structure. We have represented the latest LOINC DO structure in the MED, and developed an automated process mapping the clinical documents from both the West (Columbia) and East (Cornell) campuses to the LOINC DO. We find that the LOINC DO can represent the majority of our documents, and about half of the documents map between campuses using the LOINC DO as a reference. We evaluated the possibility of using current LOINC codes in document exchange between different institutions. While there is clear success in the ability of the LOINC DO to represent documents and facilitate exchange we find there are granularity issues.

Introduction

Clinical documents at New York Presbyterian Hospital (NYPH) are recorded in an electronic health record (EHR) system that is maintained separately for each of the two major campuses. The West Campus (Columbia University Medical Center) and East Campus (Weill Cornell Medical Center) therefore have distinct sets of clinical note types modeled in their corresponding databases which are created independently by requests from the clinicians of the two campuses. However, the clinical terms from both campuses, including documents, reside in common in the central terminology system at NYPH, the Medical Entities Dictionary (the MED)[1-3], which provides a substrate for cross-campus standardization. Currently the MED helps to manage all clinical notes created from the two campuses. With the increasing volume of notes, standardizing note names, including extension of abbreviations and conversion of local conventions, and reasonably classifying them is essential for document management and file exchange. In recent years, LOINC (Logical Observation Identifiers Names and Codes), a nationally recognized source of terminology for laboratory and other observations, has developed the LOINC Document ontology (DO) for standardizing clinical documentation with terms in a hierarchical structure. It consists of five axes: Kind of Document, Subject Matter Domain, Type of Service, Role and Setting. Each of the axes has a set of possible entry values which can be combined with values from the other axes to create LOINC-compatible combinations. A legitimate name according to LOINC DO should consist a value from "Kind of Document" and at least one value from the other four axes. In this project, all titles qualify "clinical notes" in Kind of Document axis. The most recent LOINC DO version available (at LOINC.org) had 149 possible values for subject matter, 27 for role, 18 for setting and 120 for type of service. The LOINC database additionally has included over 500 document types, basically axis combinations which have been assigned a LOINC code.

Previous work classifying NYPH document subsets according to the LOINC ontology has been reported [4-7]. The coverage of LOINC DO and codes has been evaluated by several independent studies[8-10]. In this paper we classify the entire set of documents from both campuses (a total of 2768 note titles) according to the LOINC DO in the MED (Goal 1). We then use a co-classification strategy in the MED to map common document types across both campuses to assess how this strategy could facilitate interoperability among the hospital sites with respect to notes (Goal 2). With a view toward evaluating the potential for document exchange with other institutions, we also attempted to map each campus to the set of existing LOINC codes to assess how our institutional document set matches the set that has already been assigned codes (Goal 3).

Method:

The MED has incorporated and maintained the LOINC DO for some time, but the present paper describes the recent effort to classify all documentations from both campuses within that semantic structure. The MED not only codes

the ontology values, but also maintains its hierarchy. In preparation for the mapping, we first updated the MED with LOINC DO version 2.34 (the most current at the time of the study), published on the LOINC website (at LOINC.org). A flattened representation of a small part of the Subject Matter hierarchy is seen in Figure 1.

```
Subject_Matter_Domain-94540|Emergency_Medicine-97003
Subject_Matter_Domain-94540|Emergency_Medicine-97003|Sports_Medicine-97073
Subject_Matter_Domain-94540|Emergency_Medicine-97003|Pediatric_Emergency_Medicine-97143
Subject_Matter_Domain-94540|Emergency_Medicine-97003|Medical_Toxicology-97161
Subject_Matter_Domain-94540|Emergency_Medicine-97003|Undersea_and_Hyperbaric_Medicine-97162
```

Figure 1: LOINC DO Values and Hierarchy coded in MED. Every Field contains the LOINC DO term and its unique identifier in the MED (the MED code). Every row represents parent, child, even grandchild relationships.

This project involves 1644 note titles from West Campus and 1124 from East Campus. These note titles are almost exclusively provider-authored documents related to patient care. With the total 2768 titles, an automated processing tool was developed to assign titles to LOINC DO classes in the MED. It processes every note title, extracts "Subject Matter Domain", "Type of Service", "Role" and "Setting" values from the title, and assigns it under DO classes coded in the MED. For example, note title "Psychiatry Initial Patient Evaluation – Attending" is processed and DO values are extracted: Psychiatry (Subject), Initial Evaluation (Service), Attending (Role). Since the EHR system at both sites is an inpatient clinical system, all documents, except "ambulatory", "outpatient" or specific outpatient facility specified in the title are defaulted to a "Setting" value of "Inpatient" for the purpose of mapping. The title is then assigned under "Psychiatry", "Initial Evaluation", "Attending" and "Inpatient" DO classes in the MED. Our second goal is to map East to West Campus notes using the combination of DO values. In the above example, the title is represented in the combination of DO values using MED codes. Figure 2 illustrates the process. The numbers attached to the values are MED codes.

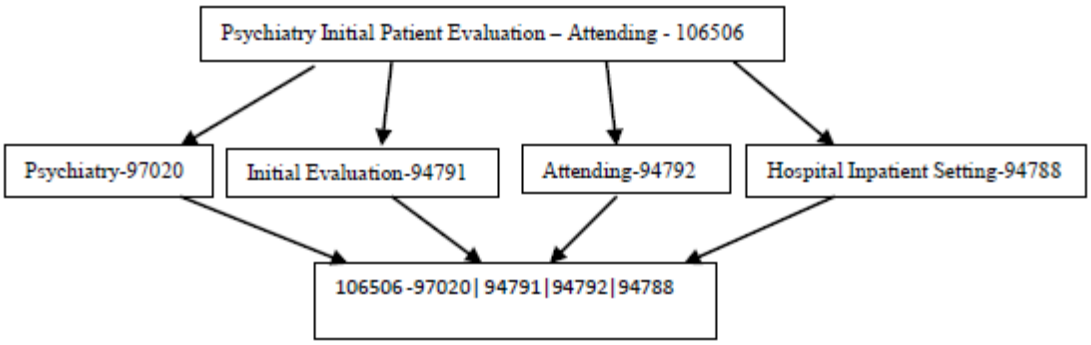

Figure 2: Processing Note Titles. Clinical note title with MED code 106506 is processed. Four LOINC DO values are extracted. The title is represented with the combination of LOINC DO values coded in MED.

Both East and West note titles are processed and coded the same way, then matched to each other. Whenever the combined values match, the two titles from the two campuses are considered mapped. With the third goal, note titles are mapped to LOINC codes. In order to use the combined DO values to do exact mapping, LOINC code descriptions are processed by our tool also. We downloaded the newest version of LOINC database, and filtered data based on scale value as "Doc" and class value as "DOC.CLINRPT" or "DOC.MISC". We also filtered out certain out-of scope codes, such as those specific for VA facilities, etc. The final data set used in this project included 374 LOINC codes. The Values from LOINC attributes "Component", "System", "Method" and "Short Name" were combined into pseudonyms for the codes, then processed by the tool to extract DO values. The codes

eventually are represented in the same format with note titles. The combined values then matched to each other. The exact match associates the local note title with the LOINC code. Figure 3 illustrates the method.

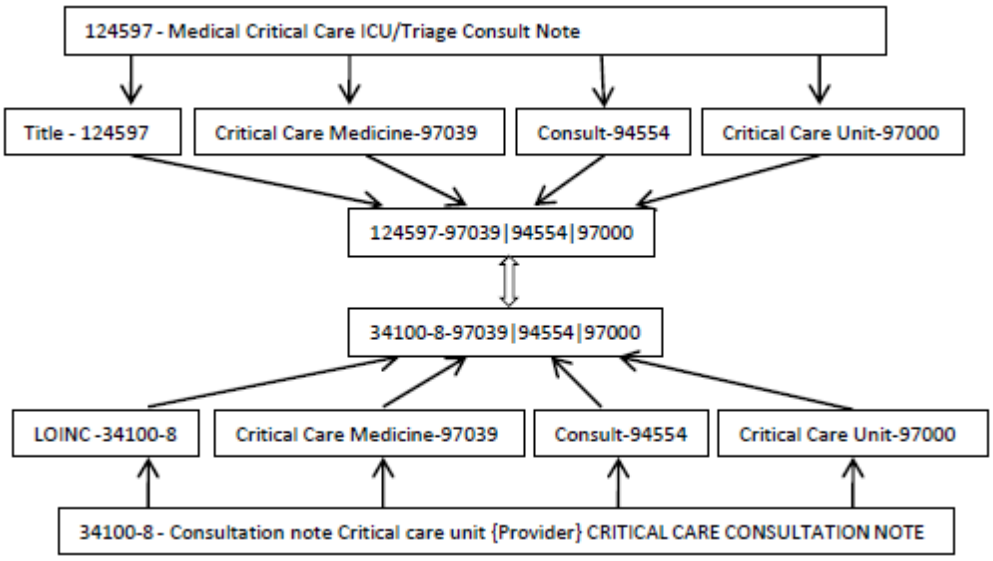

Figure 3: Mapping Note Titles with LOINC Codes. MED code 124597 and LOINC code 34100-8 are processed. LOINC DO values are extracted. The MED code and LOINC code are associated through the matching of the combinations of LOINC DO values

When doing exact matching from one campus to the other and both campuses to LOINC codes, null values are allowed in the axes values, but null must match null in the counterpart.

Results

Mapping of local documents to the individual LOINC Ontology Value Sets (Goal 1)

The majority of both East and West Campus titles map to at least one axis of the LOINC DO. Only 8.2% from West Campus and 7.1% from East Campus are left without any map to the existing document ontology (see Table 1). These unclassified items include documents such as "Pes RME Note", "Rapid Response Team Responder Note", and "Crown Clinical Snapshot", and "Eclipsys Note". Comparing the two campuses, East Campus has fewer titles than West Campus. There are notable differences in how the notes map to the LOINC DO. While both sides map at a similar rate of 68-69% to the Subject matter domain, East Campus maps to both Type of service and Role to a higher degree, and the West Campus has a higher proportion of notes where the Setting is explicitly mentioned (recall that, as described in the methods, all "Setting" values not otherwise specified are defaulted to "Inpatient" for mapping purposes since both sites are inpatient hospitals). It appears, then that the campuses have a propensity to specify their notes in a different way. For example, West Campus has a "Resident Follow up Visit Note" whereas East Campus has multiple resident follow up notes, each further indicating the Subject Matter Domain ("Cardiology Resident Follow up Note", "Dermatology Resident Follow Up Note", etc). In fact, East Campus does not include a resident follow up note title without the Subject Matter Domain included. The classification results were manually reviewed by two terminology professionals (LL and DB). Few cases were found where a dimension of a local note definition could not be matched to an axis value in the LOINC ontology. One example, "medication reconciliation", we thought could potentially be a new type of service. In a few cases, local notes specified procedure types that were more granular than those in the LOINC ontology.

Table 1. Clinical Note Title Classifications Based on LOINC DO

LOINC DO Axis	West Campus note titles (1644)	West Campus Classification Rate	East Campus note titles (1124)	East Campus Classification Rate
Subject Matter Domain	1111	67.6%	777	69.1%
Type of Service	977	59.4%	776	69.0%
Role	569	34.6%	629	56.0%
Setting (no default value)	475	28.9%	143	12.7%
Map to no axis	135	8.2%	80	7.1%

Mapping Documents Cross-Campus Using the LOINC DO Representation (Goal 2)

As can be seen from the Table 2, requiring all 4 axes to match results in 50.2% of the West Campus documents matching at least one East Campus document , and 57.2% of East Campus documents matching at least one West Campus document. In the examples shown in the two last rows of Table 2, stepwise removal of the 2 least frequently specified axes, "Setting" and "Role" raises the match rate to 71-78%.

Table 2. Mapping between West and East Campus.

Combination of LOINC DO axes (null values allowed)	West Campus note titles mapped to East Campus note titles	West Campus mapping rate	East Campus note titles mapped to West Campus note titles	East Campus mapping rate
Subject Matter Domain, Type of Service, Role, and Setting	826	50.2%	643	57.2%
Subject Matter Domain, Type of Service, and Role	953	58.0%	680	60.5%
Subject Matter Domain and Type of Service	1173	71.4%	872	77.6%

Not all notes possess four axes values, which implies that axes values are often null. For example the 2 documents "Resident Initial Visit Note" at Cornell will match to the "Resident Initial Visit Note" at Columbia although both only specify the "Role" and "Type of service".

Table 3. West and East Campus Mapping Using Non Null Values

Number of non-null axis values	West Campus (1644 total note titles)			East Campus (1124 total note titles)		
	Number of note titles found	Number mapped to East Campus note titles	Mapping rate	Number of note titles found	Number mapped to West Campus note titles	Mapping rate
4	282	135	47.9%	390	110	28.2%
3	626	239	38.2%	377	213	56.5%
2	559	379	67.8%	272	248	91.1%
1	177	174	98.3%	85	85	100.0%

In Table 3 we further evaluate the mapping rates using combinations of only NON null values. The first column represents the number of non null values each title has. The data in this table was calculated based on null values for setting defaulted to "inpatient" as described previously. 4-value means the title contains non null values from all four axes; 3-value means the title contains any combination of 3 non-null values. 2-value means any combination of two non null values, and 1-value means any single value is included. It is notable that only 282 notes are specified at all 4 axes for West Campus. This accounts for only 17 % of the total notes. Although West Campus has a greater total number of titles, East Campus has significantly greater numbers of documents specified in all four axes at 390 (35%).

Not surprisingly, the fewer highly specified documents (specified at 4 axes) on West Campus are more likely to find a match among the larger pool of specified documents on the East side (47.9%) versus 28.2% the other way.

Looking at the less-specified documents with 3 and 2 axis values, the opposite trend prevails. The less specified West Campus documents are less likely to find a match from the smaller pool of two and three axis documents on the East Campus (38 and 68% versus 57 and 91%). Documents specified at only one axis are about equivalently able to find a match on the other side (at 98-100%).

Mapping of our local documents to pre-concatenated LOINC codes (Goal 3)

Table 4 shows the results of our third goal. Notably, both East and West Campuses return matching to LOINC codes of about 20% when all four axes are included in the match. This seems low, but it is noted that many of the LOINC document codes have null values for setting and role and would therefore not match the NYPH documents that are specified for those axes. In fact, about 40% of LOINC codes have only one or two axes specified after default "Setting" values are provided, and only 10% codes have none null Subject, Service and Role values, so documents titles which have been assigned LOINC codes are, in general, less specified than those in our EHR. As predicted, removing a match at "Setting" from the requirements increases the matching rate to 23-31%, and additional removal of "Role" further increases the matching rate to 39-47%. Removing axis requirements from the matching stringency to specified LOINC codes may be appropriate in cases where LOINC has null values because it results in the matching of more specified local documents to a higher level LOINC class, For example, it results in the mapping of the 3 NYPH documents "Cardiology Fellow Consult Note", "Cardiology Resident Consult Note" and "Cardiology Attending Consult note" all to the same LOINC code 34099-2 "Cardiology Consult note".

Table 4: West and East Campus Titles map to LOINC codes

Combination of LOINC DO axes (null values allowed)	**West Campus note titles mapped to LOINC codes**	**West Campus mapping rate**	**East Campus note titles mapped to LOINC codes**	**East Campus mapping rate**
Subject Matter Domain, Type of Service, Role, and Setting	367	22.3%	210	18.7%
Subject Matter Domain, Type of Service, and Role	506	30.8%	254	22.6%
Subject Matter Domain and Type of Service	771	46.9%	441	39.2%

Discussion

We found that the LOINC DO provides a thorough semantic structure for representation of clinical documents. Fewer than 10% of NYPH documents had no assignment to any axis. Many of the non-mapped notes appeared to be either items specific to our local environment (such as "Crown Clinical Snapshot"), abbreviated names that were not clearly evident ("Pes RME Note"), and possibly due to missing DO values ("Rapid Response Team Responder Note").

Although the LOINC DO covers most of our local term needs, we do note a few limitations. For example, the Type of service "Consult Follow-Up" seems to be missing in the version we used. This forces us to choose Is it consult or follow up. We could take advantage of the poly hierarchy of the MED to place it under both, but then that introduces ambiguity when mapping. There are a few potential ambiguities in LOINC DO terms and classifications. For example, In LOINC, "Neurology Surgery", "Plastic Surgery" and "Orthopedic Surgery" have parallel relationships with Surgery, not IS-A relationship. "Surgery of the hand" appears three times as a child under "Surgery", "Plastic Surgery" and "Orthopedic Surgery". They may mean surgeries with different purposes, but giving the same name with current hierarchy is confusing.

It was initially surprising that the campus site with the greater number of clinical note titles tended toward lower specification (map to fewer LOINC axes). A follow up review of the documents at that campus showed that they model some documents from a second hospital site and also tend to have more ambulatory settings included (which accounts for part of the higher mapping to the "Setting" axis for West Campus in Table 1).We expect that this classification scheme will have real life impact on helping clinical applications display notes in a standard manner, perhaps using the multiple hierarchy in the MED to more quickly find notes for entry or reviewal.

One of the objectives of this project was to test the use of ontology – modeled notes for document interoperability between campuses. For such a use case, the issue of granularity came into play. This has been noted by others as well [5, 7, 9]. With one campus apparently naming their notes more specifically (therefore mapping to more LOINC axes per note), equivalency of documents on the two campuses is more difficult to establish. This could in part reflect how the documents are actually being used. For example, if most West Campus residents use the same note title "Resident Follow up Visit Note" for their follow up notes, whereas East Campus residents use one of 26 different titles which further specify the Subject Matter, as we described in the results. But, in some cases, the distinction in granularity might simple reflects naming conventions. These two possibilities may be hard to distinguish. That being said, we were able to exactly map over 50% of the documents on each campus to a document on the other.

Addressing the possibility of using LOINC codes as a bridge between local documents and outside institutions yielded the finding that a relatively small percentage of our local documents find exact matches to existing LOINC codes (18-22%). Further examination showed that there are only 37 out of 374 (10%) LOINC codes with NON null values for Subject, Service and Role. This indicates that most of the matching documents only matched at one or two axes. So the granularity of specification intercedes in precise matching to LOINC codes, This implies that, at least at the current stage, using these codes for exact document exchange may not be feasible, however, the specification of classes of documents certainly is. Specific decision can be made to omit axes from matching criteria. For example, cardiology consult notes can be represented even if specification of those consults by role do not yet have distinct LOINC codes. Of course, LOINC codes are dynamic, and new codes can be requested which have required specification. In addition, even if specific codes are not yet created, the LOINC DO can represent documents at the more granular level.

Conclusion

The LOINC document ontology model can provide a robust representation of clinical documents. This can be exploited to assist in common naming conventions, finding notes by traversing the axis categories of Subject, Setting, Role and Type of Service. However, the granularity of representation can be an issue in mapping across sites using the LOINC DO and mapping to specific LOINC codes.

References

1. Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. J Am Med Inform Assoc. 2000 May-Jun;7(3):288-97.
2. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998 Nov;37(4-5):394-403.
3. Cieslowski BJ, Wajngurt D, Cimino JJ, Bakken S. Integration of nursing assessment concepts into the medical entities dictionary using the LOINC semantic structure as a terminology model. Proc AMIA Symp. 2001:115-9.
4. Bakken S, Cimino JJ, Haskell R, et al. Evaluation of the clinical LOINC (Logical Observation Identifiers, Names, and Codes) semantic structure as a terminology model for standardized assessment measures. J Am Med Inform Assoc. 2000 Nov-Dec;7(6):529-38.
5. Hyun S, Bakken S. Toward the creation of an ontology for nursing document sections: mapping section names to the LOINC semantic model. AMIA Annu Symp Proc. 2006:364-8.
6. 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.
7. 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.
8. Brown SH, Lincoln M, Hardenbrook S, et al. Derivation and evaluation of a document-naming nomenclature. J Am Med Inform Assoc. 2001 Jul-Aug;8(4):379-90.

9. Chen ES, Melton GB, Engelstad ME, Sarkar IN. Standardizing Clinical Document Names Using the HL7/LOINC Document Ontology and LOINC Codes. AMIA Annu Symp Proc.2010:101-5.
10. 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.