MAKING THE "MINIMUM DATA SET" COMPLIANT WITH HEALTH INFORMATION TECHNOLOGY STANDARDS

ATTACHMENT B:

ENCODING NURSING HOME RESIDENT MDS OBSERVATION AND ASSESSMENT DATA ARTICLE

Encoding Nursing Home Resident MDS Observation and Assessment Data: Do HL7 Messaging Standards Support its Transmission?

1. Background

Apelon, Inc. has been tasked by the U.S. Dept. of Health and Human Services (HHS) to explore standard terminology representations of the content in its "Minimum Data Set (MDS) Version 2.0 for Nursing Home Resident Assessment and Care Screening" survey instrument. SNOMED-CT[1] has been proposed as a candidate standard terminology that can capture and encode most of the clinical concepts expressed in the questions and responses from the MDS survey. Often, however, several SNOMED-CT concepts must be combined or post-coordinated to represent more complex MDS meanings accurately. SNOMED-CT appears to have many of the required "nouns" or "adjectives" (concepts, qualifiers, and modifiers), and "verbs" or "linkage" concepts (associations, role-relationships) to construct these complex expressions.

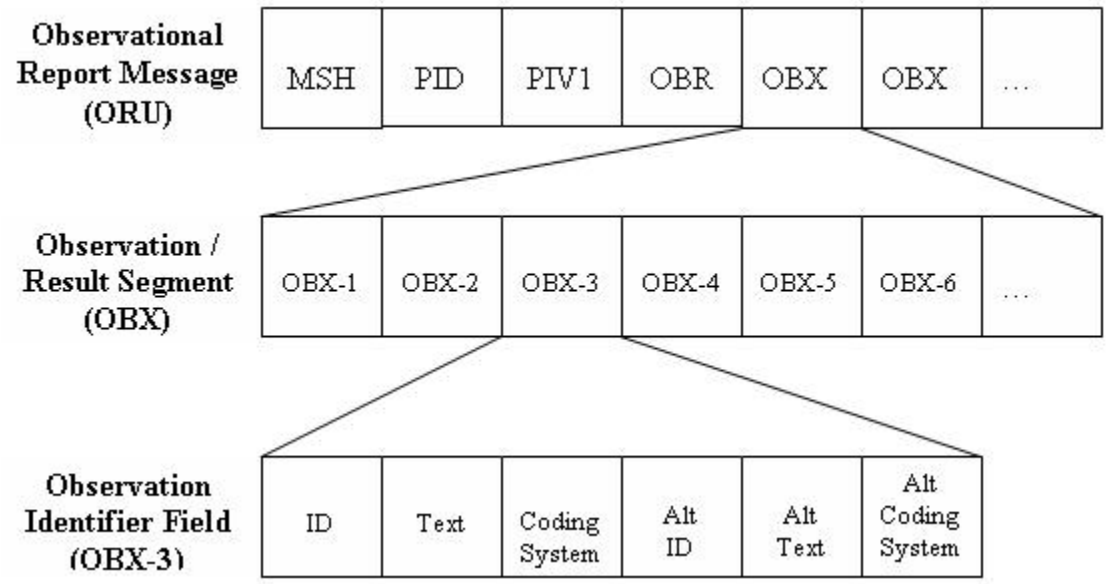
If post-coordinated terminological expressions are indeed required to encode and to adequately represent portions of the MDS content, can such encoded data then be communicated electronically from nursing home systems to the HHS by accepted Health Level 7 (HL7) standard messages? Does Version 2.5 of the HL7 Standard for electronic data exchange in healthcare environments,[2] the latest ANSI-standard release, support transmission of these messages? Would enhancements proposed in the HL7 Version 3.0 draft standard provide a better solution? This document researches answers to these questions.

2. HL7 Version 2.5 Message Standards

HL7 Version 2.5 specifies an extensive collection of standard messages and exchange protocols for electronic healthcare data exchange. Messages consist of a group of required or optional message segments in a defined sequence, which together conveys specific types of information (admission/ discharge/transfer, financials, pharmacy reporting, orders, observations, and so on). Certain message segments can be reused in many different types of messages to transmit data for a particular domain (e.g., results of observations). Message segments themselves contain logical groupings of required or optional data fields, which are delimited strings of characters, constrained as pre-defined datatypes. Figure 1 illustrates some of these building blocks for a HL7 Version 2.5 message for observational reporting.

Datatypes relevant for the transmission of coded data (generic or post-coordinated) are the Coded Element (CE) datatype and its specializations: Coded with No Exceptions (CNE) and Coded with Exceptions (CWE). The latter two differentiate whether the set of available codes for a particular sub-domain is a closed world or open to local additions. For example, the OBX-3 field in Figure 1 is a CE datatype.

FIGURE 1. Components of a HL7 Version 2.5 Message for Observational Reporting



The CE and related datatypes consist of up to 6 components, delimited by '^' in the actual message stream. The following table describes these components in more detail.

CE Datatype Component	CE Component Description	Sample Value
Identifier	Specific code from coding system	44950
Text	Text equivalent of code	Appendectomy
Name of Coding System	HL7 ID of coding system[I]	CPT
Alternate Identifier	Alternate code from other coding system	P1-57450
Alternate Text	Text equivalent of alternate code	Appendectomy, NOS
Name of Alternate Coding System	HL7 ID of alternate coding system	SNM

For example, to encode an appendectomy using CPT in an HL7 Version 2.5 message, one would use a **Coded Element (CE)** datatype to transmit the proper CPT code for an appendectomy, its text name, and an HL7-designated ID for the CPT coding system. Generally, only these first three components are required, and some of them may be optional in selected message field usage.[3] The last column of the above table also demonstrates how the final three components would be used to transmit a synonymous code, here mapping the primary code from CPT to an alternate code in legacy SNOMED International.

2.1 Observation/Result Segment (OBX)

Although structured, patient-oriented clinical data can be transmitted within several HL7 message types for different purposes (medical document management, results reporting, clinical trials, etc.); all such message types use the Observation/Result Segment (OBX) to transmit *each individual clinical observation*, the smallest indivisible unit of a report. Numerous OBX segments must be assembled to send all the observations in a report. They can be bundled into explicit panels or batteries of observations (e.g., electrolytes, vital signs, or sections of a survey) by following a shared header, an Observation Request Segment (OBR). Clinical data to be sent via OBX segments include, but are not limited to, patient history and physical, consultations, operative reports, discharge summaries, pathology reports, imaging reports, laboratory results, waveform results (EKG), and survey results.

The OBX segment contains up to 19 fields (full discussion beyond the scope of this document). Most relevant to the encoding of MDS content are the three OBX fields listed next, which are numbered by their position in that message segment:

Field Number	Field Name	Field Description
OBX-2	Value Type	Datatype of the result value in Field #5
OBX-3	Observation Identifier	ID or code naming the observation
OBX-5	Observation Value	Observed value for named observation

The OBX-3 Observation Identifier encodes the name of the thing observed (e.g., Serum sodium, Diastolic BP, Pain Site, etc.) via a *single* code from some master observation table or an external coding system, such as LOINC.[4] The use of identifiers from external, authoritative coding systems or terminologies improves interoperability across healthcare sites. OBX-3 is always an instance of the CE (coded element or entry) datatype.

Although a limited set of identifier code suffixes (e.g. &IMP for diagnostic impression, &ANT for anatomy, and so on) have been defined by HL7 for post-coordination with OBX-3 codes, the OBX-3 field *does not* allow the ID of an observation (e.g., an MDS survey question) to be an arbitrary, post-coordinated expression of codes. Complex concepts must therefore be pre-coordinated, single entities in the master observation table or external coding system, per HL7 Version 2.5 standards.[5]

The OBX-5 field contains the resulting *value* for that observation identified by OBX-3. Although OBX-5 values can be transmitted as any suitable datatype, HHS MDS project requirements mandate that they be coded data. Encoded OBX-5 observation values must be flagged with a "CE" value, denoting a coded element or entry, in the OBX-2 value type field. HL7 allows code values from multiple external coding systems, including all HIPAA and CHI designated terminologies. Since each logically-independent observation must be reported in a separate OBX segment, batteries or panels of observations (e.g., Vital signs) would consist of multiple OBX segments.

However, the HL7 Version 2.5 specification for OBX-5 *does* permit post-coordination of codes that together describe a modified or qualified value for a *single, logically independent* observation result value. It does so by permitting *repeating* OBX-5 fields within a *single* OBX message.

> Though two independent diagnostic statements cannot be reported in one OBX segment, multiple categorical responses are allowed (usually as CE datatypes separated by repeat delimiters), so long as they are fragments (modifiers) that together construct one diagnostic statement. Right upper lobe (recorded as one code) and pneumonia (recorded as another code), for example, could be both reported in one OBX segment. Such multiple "values" would be separated by repeat delimiters.[6]

Although the HL7 Version 2.5 standard does not specify a grammar for associating or combining repeating codes, certain healthcare terminologies (e.g., SNOMED-CT) provide linkage concepts in order to build post-coordinated expressions. Receiving applications must be able to process these messages and be programmed to assemble sequential codes correctly.

The next example, taken from the HL7 documentation, illustrates an HL7 Version 2.5 OBX message stream which incorporates both an encoded, post-coordinated observation value (OBX-5), as well as a suffix tag to modify the meaning of the observation identifier (OBX-3). Message segment fields are delimited by the '|' character. The OBX segment fields of interest are highlighted.

```
OBX|1|CE|24646-2&IMP^CXR PA+LAT^LN||
       .61^RUL^ACR~.212^Bronchopneumonia^ACR|||A|||F|...
```

A LOINC code (24646-2) serves as the OBX-3 observation identifier, encoding the observation name as a diagnostic impression of a PA and lateral chest X-ray. The OBX-3 CE field consists of 3 (^ delimited) components: the 1st (ID) containing a hyphenated LOINC code plus an IMP[ression] suffix tag, the 2nd

(text) displaying an abbreviated LOINC name for this chest X-ray, while the 3rd (coding system ID) designates LOINC itself. The OBX-5 observation value field, a coded element according to OBX-2, contains two post-coordinated codes, both from the ACR (American College of Radiology) coding system, which are separated by a field repeat (~) delimiter. Post-coordination permits encoding of the concept "Right Upper Lobe Bronchopneumonia" via a simple serialization of ACR codes .61 and .212

To reiterate a key issue for MDS encoding, although the OBX-5 observation value permits repeating occurrences to build a fully-specified, post-coordinated result, the OBX-3 field does not. Each OBX-3 observation identifier must be a single code, uniquely identified in some HL7 or local master table, or an authoritative external coding system.

The LOINC coding system has become the *de facto* external standard system for OBX-3 identifier codes, such that several organizations have submitted their nursing surveys for inclusion in and encoding by LOINC. Items from these nursing instruments are labeled as instances of the "Survey" classtype with a suitable LOINC class value, an officially-assigned LOINC code, and an optional list of valid answers. The latest LOINC Version 2.15 has the following short list of classes within the Survey classtype, each containing numerous LOINC-encoded items:

LOINC Class	Description
SURVEY.NURSE.HHCC	Home Health Care Classification Survey
SURVEY.NURSE.HIV-SSC	Signs and Symptoms Checklist for Persons with HIV Survey
SURVEY.NURSE.LIV-HIV	Living with HIV Survey
SURVEY.NURSE.OMAHA	OMAHA Survey
SURVEY.NURSE.QAM	Quality Audit Marker Survey

For example, the following table shows several LOINC-encoded items from the Home Health Care Classification (HHCC) survey instrument. Each item happens to have the same ANSWERLIST attribute value, the set of valid values expected for results, namely: IMPROVED, STABILIZED, DETERIORATED.

LOINC Code	Fully-specific LOINC Name
28079-2	ACTIVITIES OF DAILY LIVING ALTERATION: FIND:PT:^PATIENT:ORD:OBSERVED.HHCC
28080-0	ACTIVITY ALTERATION:FIND:PT:^PATIENT:ORD:OBSERVED.HHCC
28081-8	ACTIVITY INTOLERANCE:FIND:PT:^PATIENT:ORD:OBSERVED.HHCC
28082-6	ACTIVITY INTOLERANCE RISK:FIND:PT:^PATIENT:ORD:OBSERVED.HHCC
28083-4	ACUTE PAIN:FIND:PT:^PATIENT:ORD:OBSERVED.HHCC
28191-5	POISONING RISK:FIND:PT:^PATIENT:ORD:OBSERVED.HHCC

Constrained by limitations of the current HL7 Version 2.5 standard, HHS could designate its MDS survey instrument items as coded value sets for a master HL7 observation table or for incorporation in future versions of the LOINC coding system. In either case, each MDS survey item would receive a unique code suitable for the OBX-3 *observation identifier* field.

MDS survey items and their codes could also be assigned officially-maintained external mappings to post-coordinated SNOMED-CT expressions, which then define and deconstruct their meanings in terms of more atomic SNOMED-CT reference terminology concepts. Although not transmitted as part of an HL7 message, those mappings should be published and could be used by recipient and HHS systems for data analytic and aggregation purposes.

Since the OBX-5 *observation value* field already permits post-coordinated codes, a SNOMED-CT-based solution for results reporting can already be provided with the current HL7 Version 2.5 messaging standard.

2.2 Clinical Data Architecture (CDA) Document

The HL7 Clinical Data Architecture (CDA) Release 1.0[7] is a document markup standard that specifies the structure and semantics of clinical documents for exchange purposes. It subdivides documents into meaningful, tagged chunks of information and provides a template for structuring computably-valid instances of a clinical document. Although derived from early versions of HL7 Version 3 Reference Information Model (RIM) and Abstract Data Types *draft* standards, CDA has already achieved ANSI-standard certification. Some of its constructs transition between the two HL7 versions.

Clinical CDA documents are complete information objects encoded in E**X**tensible **M**arkup **L**anguage (XML)[8] and may include multimedia content. At the present time, they can be a MIME-encoded payload *within an HL7 Version 2.5 message*. For example, an HL7 Version 2.5 OBX observation segment can contain a complete CDA document as the OBX-5 observation value, flagged by an OBX-2 value of "ED" (Encapsulated Data) and by other means.

At the risk of simplifying the CDA too greatly, the <body> of a CDA document consists of nested <section>, <paragraph>, <list>, <item>, <table> and/or other XML markup elements, as specified by a formal CDA document type description (DTD) developed by HL7. <content> and <coded_entry> elements can be used to markup and encode clinical content from a variety of domains. The <coded_entry> element inserts codes from HL7-recognized coding schemes into CDA documents. A <coded_entry.value> element can explicitly reference the original text within the document that is being encoded.

Vocabulary domains provide the value sets for CDA-required coded attributes, as well as optional <coded_entry> elements. Value sets can be HL7-specified concepts or defined subsets of recognized external coding systems such as LOINC or SNOMED-CT. HL7 assigns a unique identifier to each vocabulary domain, and every concept within such a domain must have a unique code.

The following example, taken from the CDA specification, illustrates concept coding in a CDA document. A sample problem-oriented medical record section has a <caption_cd> element, which provides the LOINC code (V=code value S=coding system ID) for the <caption> element value "Assessment". The Assessment record consists of a <list> of three <item> elements, but only the first has coded <content>. A <coded_entry> element provides the SNOMED International code for "Asthma", text marked up by the previous <content> element which assigned it an internal ID="String001".

```
<section>
    <caption>
        <caption_cd V="11496-7" S="2.16.840.1.113883.6.1"/>Assessment
    </caption>
    <list>
        <item>
            <content>
                <content ID="String001">Asthma</content>, with prior smoking
                        history. Difficulty weaning off steroids. Will try
gradual taper.
                <coded_entry>
                    <coded_entry.value ORIGTXT="String001"
                         V="D2-51000" S="2.16.840.1.113883.6.5"/>
                    </coded_entry>
            </content>
        </item>
        <item><content>Hypertension, well-controlled.</content></item>
        <item><content>Contact dermatitis on finger.</content></item>
    </list>
</section>
```

The CDA framework permits multiple <coded_entry> elements for <content>, with the original text marked up or not. It is our understanding that relevant HL7 Version 3 coded datatypes will also permit post-coordinated codes.[9]

Extrapolating from this example, the MDS survey could be represented as a collection of tables. An MDS question, as the <content> element in the first column of such a table, might use a LOINC <coded_entry> element to encode it in its entirety, with portions of its <content> optionally marked up and further encoded by SNOMED-CT. The MDS *answers* or *results*, as a <list> of <item(s)> in the second column of the CDA table, would have its <content> marked up and encoded by SNOMED-CT <coded_entry> elements.

3. HL7 Version 3.0 Message Standards

HL7 Version 3.0 remains a *draft* standard at this writing. Unlike HL7 Versions 2.x, which have evolved for more than a decade via a "bottom-up" approach permitting extensive “optionality”, Version 3.0 is being developed using a Reference Information Model (RIM) for data and a “top-down” object-oriented methodology to create concise, testable, well-defined messages. The RIM provides an explicit representation of the semantic and lexical connections existing between the information to be carried in implemented HL7 messages. Due to its complexity and on-going evolution, even a cursory overview of HL7 Version 3.0 is well beyond the scope of this document. HL7 members can obtain the latest draft versions of the HL7 Version 3.0 RIM, the Abstract Data Type Specification, and the Data Type Implementable Technology Specification for XML at Members Only portions of HL7’s official web site (<http://www.hl7.org>)

HL7 Version 3.0, when approved and fully-implemented, should be capable of meeting HHS MDS encoding and messaging requirements. Using the Concept Descriptor (CD) datatype, it will permit post-coordinated encoding of both observation identifiers and actual values. Detailed rules and policies for post-coordination semantics are still under development by HL7 committees. Nevertheless, one can glimpse how the HL7 Version 3.0 RIM and concept descriptor datatype will address post-coordinated encodings.

HL7 committees and other organizations have formalized the notion of a version 3.0 “Clinical Statement” as an expression of a discrete item of clinical information and its context, as relevant to a specific patient. Clinical statement patterns provide a common structural framework and model, derived from specified classes and attributes in the RIM (e.g., Act, Encounter, Observation, etc.) and connected by Statement Relationship linkages, to express detailed clinical content. Encoded concepts for content in HL7 clinical statements can be taken from a clinical reference terminology such as SNOMED-CT, which permits complex, post-coordinated expressions.

Clinical content, including observation identifiers and observation values, can be transmitted as instances of the Concept Descriptor (CD) abstract datatype in HL7 Version 3.0 messages. Like Version 2.5 legacy CE datatypes, the CD datatype can transmit a code (e.g., a SNOMED-CT conceptID), the name of the coding scheme for that code, a display name for the code, and optional synonyms, as well as the original text being encoded. The Version 3.0 CD datatype builds on the CE with a grammar for post-coordinating codes from a terminology to create a new concept. The Concept Descriptor grammar allows the assignment of modifiers, specifically: named roles and their values, where values themselves can be further modified.

The following SNOMED-supplied examples illustrate how two closely related surgical procedure concepts can be transmitted in HL7 Version 3.0 XML <Observation> messages using CD datatyped elements to express post-coordinated SNOMED-CT encodings.

In the first example, the <Clinical_procedure> class element has an <approach_site_cd> element “transfrontal approach” which is logically AND’ed[10] with the primary concept descriptor <cd> element

"hypophysectomy" to express the desired post-coordination. The original concept "hypophysectomy by transfrontal approach" thereby encoded is also captured by a <txt> element.

```
<Observation>
    <!-- other elements here -->
    <Clinical_procedure>
        <cd code="52699005" displayName="Hypophysectomy" codeSystem="SNOMED
CT"/>
        <txt>Hypophysectomy by transfrontal approach</txt>
        <approach_site_cd code="65519007"
           displayName="Transfrontal approach" codeSystem="SNOMED CT"/>
    </Clinical_procedure>
</Observation>
```

The next example shows another method of post-coordination supported only by the full CD datatype. The primary concept descriptor <cd> element itself has been further modified with a SNOMED-coded <modifier> role name and value, thereby altering that <cd> to become an "incision of a brain lesion", where "lesion" is the value of the "DIRECT-MORPHOLOGY" role modifying "brain incision". SNOMED-CT concept representation semantics guide the choice of roles and values used to construct post-coordinated concepts. Also, as in the previous example, the <Clinical_procedure> message specification allows an additional <target_site_cd> element "pituitary posterior lobe" which is also logically AND'ed with the modified primary concept descriptor <cd> element to express the fully post-coordinated, encoded concept. Here again, the original concept is captured by the <txt> element "incision of lesion of posterior lobe of pituitary gland".

```
<Observation>
    <!-- other elements here -->
    <Clinical_procedure>
        <cd code="42699003" displayName="Brain incision" codeSystem="SNOMED
CT">
            <modifier code="49755003" displayName="lesion"
codeSystem="SNOMED CT">
               <name code="363700003" displayName="DIRECT-MORPHOLOGY"
                  codeSystem="SNOMED CT"/>
            </modifier>
        </cd>
        <txt>Incision of lesion of posterior lobe of pituitary gland</txt>
        <target_site_cd code="37512009"
           displayName="Pituitary posterior lobe" codeSystem="SNOMED CT"/>
    </Clinical_procedure>
</Observation>
```

By analogy, one can envision how MDS nursing survey observations and results could also be transmitted as HL7 <observation> or other clinical statement messages via post-coordinated encodings of SNOMED-CT through concept descriptor datatypes. These examples also suggest the need for HL7 efforts that are still on-going to clarify and specify how to ensure unambiguous semantics for post-coordinated encoding.

References

[1] *SNOMED Clinical Terms*. SNOMED International, July 2005. <http://www.snomed.org/snomedct/>.

[2] *Health Level Seven Version 2.5 Final Standard*, July 2003. <http://www.hl7.org>.

[3] In certain instances, a CE datatype can even transmit uncoded free-text data. However, our focus is on the transmission of coded data from standard terminologies.

[4] *Logical Observation Identifiers Names and Codes Version 2.15*. Regenstrief Institute, Indianapolis, IN, June 2005. <http://www.regenstrief.org/loinc/>.

[5] Personal communications with Drs. Clem McDonald and Stan Huff, August 2005.

[6] *Health Level Seven Version 2.5 Final Standard*. July 2003, pp. 7-47. <http://www.hl7.org>.

[7] *Health Level Seven Version 3 Standard: Clinical Data Architecture Release 1.0*. November 2000. <http://www.hl7.org>.

[8] *Extensible Markup Language (XML)*. World Wide Web Consortium. <http://www.w3.org/XML/>.

[9] Personal communication with Dr. Stan Huff, August 2005.

[10] HL7 Version 3 RIM-based specification of the <Clinical_procedure> element defines a semantically-suitable set of optional concept descriptor elements for procedures, such as approach site or target site, to be logically AND'ed with the primary concept descriptor.

MAKING THE "MINIMUM DATA SET" COMPLIANT WITH HEALTH INFORMATION TECHNOLOGY STANDARDS

PDF Files Available for This Report

Main Report http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT.pdf

ATTACHMENT A: BIPA, Sec. 545. Development of Patient Assessment Instruments
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-A.pdf

ATTACHMENT B: Encoding Nursing Home Resident MDS Observation and Assessment Data: Do HL7 Messaging Standards Support its Transmission?
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-B.pdf

ATTACHMENT C: Side-by-side depiction of MDSv2 and Content Matching Results
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-C.pdf

ATTACHMENT D: Additional Item Matching (emerging MDSv3 items)
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-D.pdf

ATTACHMENT E: NLM/UMLS to Maintain Links between LOINC coded Assessment Question and Answers and Codeable Vocabularies -- An Alternative
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-E.pdf

ATTACHMENT F: The specific vocabulary codes that correspond with each of the standardized vocabulary matches identified in Attachments C and D
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-F.pdf
This Attachment is also available as an **Excel** file at:
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT-F.xls

This full report and links are also available in **HTML** format at:
http://aspe.hhs.gov/daltcp/reports/2006/MDS-HIT.htm