

An Alternative CDISC-Submission Domain for Laboratory Data (LB) for Use with Electronic Health Record Data

Jozef Aerts¹

¹ Institute for eHealth, University of Applied Sciences FH Joanneum, Graz, Austria

Abstract

Background: The CDISC SDTM standard for submission of clinical study data to the FDA was developed at a time when the extraction of data from electronic health records or hospital information systems was still uncommon. Therefore the current SDTM is not well suited for cases where interoperability between healthcare and research has already been realized.

Objectives: It is therefore necessary to adapt the SDTM to accommodate for these present-day use cases.

Methods: A critical analysis of the existing "Laboratory" (LB) SDTM domain has been made with respect to the suitability to represent data extracted from electronic health records.

Results: An alternative "Laboratory" domain (abbreviated LN – Laboratory New) for usage with data from electronic health records is presented.

Conclusions: The alternative LN domain presented fulfills the requirements for direct population with data from electronic health records. As a by-product, it allows reviewers at the FDA to actually compare laboratory data between studies and submissions which was not possible with the classic SDTM "Laboratory" domain.

Keywords

Interoperability, semantics, electronic health records, regulatory submissions, FDA, CDISC, SDTM, Laboratory domain, LOINC, UCUM

Correspondence to:

Jozef Aerts

Institute for eHealth, University of Applied Sciences FH Joanneum
Address: Eggenberger Allee 11, A-8020 Graz, Austria
E-mail: Jozef.Aerts@fh-joanneum.at

EJBI 2015; 11(1):en2-en9

received: February 2, 2015

accepted: June 1, 2015

published: June 30, 2015

1 Introduction and Background

Considerable progress has been made in the last few years as to the integration of clinical research data capture and electronic health records (EHRs). The IHE Profile "Retrieve Form for Data Capture" (RFD) [1] has created the technical framework for retrieving information from EHRs and automatically prefilling clinical case report forms with data from the EHR.

At the other end of the spectrum, for electronic submissions, the FDA strongly encourages [2] the use of the "Study Data Tabulation Model" (SDTM) from the CDISC (Clinical Data Interchange Standards Consortium) organization [3], a semantic standard requiring captured data to be categorized, rearranged, combined, or to be derived, and to be listed in tables. The SDTM standard is regularly updated (the most recent version being 1.4). Upon each new version, new domains and new variables are added, based on the needs of the FDA, and in rare cases, some are removed. This means that the "Implementation Guide" (SDTM-IG) is growing in size with each new release, and the standard is becoming more complex [4].

Little attention has as yet been paid to the question whether the SDTM is "EHR friendly", i.e. whether data coming from electronic health records can be easily used in SDTM tables without the need of complex transformations that can lead to errors or information loss. This is important, as due to continuously improving technical integration, data is coming more and more frequently from the EHR instead of being manually captured by the investigator.

As an example, we investigated in how much the SDTM "Laboratory" (LB) domain is fit for integration with semantic data standards used in healthcare and in EHRs in particular.

1.1 Semantic Standards for Laboratory Data used in Healthcare

There are two important semantic standards used in healthcare in the area of laboratory data. The first is LOINC [5] from the Regenstrief Institute which is a coding system for laboratory tests. The latest release (2.50) contains over 72,000 codes, both for laboratory tests and for vital signs tests as well as document related codes.

However, the majority of the test codes are laboratory test codes. Each test code consists of a 3 to 5 character digit, followed by a dash, and a 1 character check digit. LOINC also published a list of "Top 2000+ test codes" which accounts for over 98% of the volume of tests in hospitals and central labs [6].

LOINC is not just a list of test codes, it is a system. Essentially, it is a 5-dimensional system with a 6th optional dimension describing the test method when necessary. Each LOINC test code is described by 5-6 variables, which is known as a "LOINC fully specified name". So each LOINC term depicts the following structure:

```
<component/analyte>:<kind of property>:<time aspect>:<system type>:<scale>:<method>
```

An example is "Glucose:MCnc:24H:Urine:Qn", meaning: Glucose measured as mass concentration (MCnc) in 24 hour urine, quantitative, and having test code 21305-8.

Another example is "Glucose:MCnc:Pt:Urine:Qn" meaning: Glucose (analyte) measured as mass concentration (MCnc) as a point in time (Pt) in Urine, quantitative (Qn). The test code for this test is 2350-7. The code itself does not contain any meaning.

Although looking very similar, the expectation values and normal ranges can be totally different.

As each of the 5 (or 6) parts follows controlled terminology, the number of combinations can be extremely large (however not infinite) but not each combination will have a code. This is important for the understanding of this article.

The Regenstrief Institute also developed and released a computer software (RELMA) [7] to search a LOINC database and to develop mappings to specific local vocabularies as have been developed in many hospitals.

The second important semantic standard that is used in healthcare and of special importance for lab data and physical quantities in general is UCUM (Unified Code for Units of Measurement) [8]. Like LOINC, it is not a list but a system. Even more, it contains a set of rules on how unit codes can be generated. A UCUM unit of measure usually consists of two or three parts, namely a prefix (like "m" for "milli"), a base unit (e.g. "m" for "meter") and possibly a further designator, like [Hg] for "mercury column". The combination "mm[Hg]" then designates "millimeter mercury column" which is a unit for the property "pressure". In addition, an XML file containing all the prefixes, base units, designators, and also special (non-SI) units (like [in_i] for inches), the "ucum-essence.xml" file [9], allows to generate software for automated conversion between units for the same property.

It is important to note here that the use of UCUM is mandatory in both HL7-v3 [10] and in ISO-21090 data types [11] when the object is of type "PQ" (physical quantity). So in EHR exports, any data point that corresponds to a physical property will have a UCUM unit, with the exception of physical properties that do not have a unit, like a pH.

Also note here that the CDISC Operational Data Model standard (ODM), the worldwide standard for exchange and archival of data in clinical research, is already able to take HL7-v3 data points as well as ISO-21090 formatted data points (or even HL7-FHIR resources) [12]. This may be of importance when discussing an XML based exchange format for submission data sets that also can include data points from electronic health records.

2 Methods: Analysis: The SDTM Laboratory (LB) Domain – Current Situation

The LB domain in the SDTM-IG describes a dataset as a table with a number of variables like STUDYID (study identifier), USUBJID (unique subject identifier), LBTESTCD (lab test code), LBTEST (lab test name – 1:1 relation with LBTESTCD), LBORRES (original result), LBORRESU (original units) etc.. Some of these variables have originally been directly captured using a CRF, others are assigned (such as LBSEQ – sequence number, and LBBLF – baseline flag), and again others are clearly derived, such as "LBDY" (Study day of Specimen Collection). Table 1 gives a selection of the variables that we will discuss further on, and for which we want to propose alternatives. This table also contains a column stating whether the variable value is governed by controlled terminology, i.e. whether the values are restricted to be one of controlled terminology terms published by CDISC, and the name of the controlled terminology list in square brackets.

SDTM uses a surrogate key in most domains which is LBSEQ (sequence number) in this case: it is unique within each subject in the table, i.e. the combination of STUDYID (which has a fixed value within the table), USUBJID (subject ID) and LBSEQ (sequence number) forms the primary key of the table, although SDTM is not a relational database, but more a "view" on a database.

As such, the value of LBSEQ is usually assigned in the very last step of the table generation.

The assignment of natural keys is case dependent and is performed by the sponsor, and documented in the meta-data file (the "define.xml" file). Usually, for the LB domain, the natural keys (or key candidates) are STUDYID, USUBJID, LBTESTCD, LBSPEC (specimen type), LBMETHOD (method of test or examination), VISIT or VISITNUM (visit name and number – also a 1:1 relation), LBDMTC (date/time of collection) and/or LBTPPT or LBTPPTNUM (planned time point name and/or number). Also LBLOINC (LOINC code) can be a candidate key. The interesting fact in the Implementation Guide is that LBLOINC is described as "dictionary-derived code for LBTEST". The wording "derived" implies that commonly, the laboratory does not provide the LOINC code together with the test results. Instead, it needs to be derived from the other available information, although the executing laboratory very probably used it internally. Due

Table 1: Most important variables in the SDTM "LB" domain.

Variable	Description	Required / Expected / Permissible	Controlled Terminology [name]
USUBJID	Unique Subject ID	Required	
LBGRPID	Group ID – used to tie together related records	Permissible	
LBREFID	Specimen ID	Permissible	
LBTESTCD	Test Code	Required	Yes [LBTESTCD]
LBTEST	Test Name (1:1 relationship with LBTESTCD)	Required	Yes [LBTEST]
LBCAT	Test Category – e.g. HEMATOLOGY	Expected	No
LBSCAT	Subcategory	Permissible	No
LBORRES	Original result	Expected	
LBORRESU	Original result units	Expected	Yes [UNIT]
LBSPEC	Specimen type	Permissible	No
LBMETHOD	Method of test or examination	Permissible	No

Table 2: Example laboratory test variables from different submissions of different sponsors.

Sponsor	LBTESTCD	LBCAT	LBSPEC	LBMETHOD
Sponsor 1	GLUC	CHEMISTRY	BLOOD	QUANT
Sponsor 2	GLUC	CHEM	WHOLE BLOOD	ENZYMATIC
Sponsor 3	GLUC	CHEMI	BLOOD	HEXOKINASE

to this and the fact that LBLOINC is stated as being "permissible", very few submissions contain the LOINC codes for the tests. Asked why the laboratories do not provide the LOINC codes to the investigators, a commonly received answer was like "we do use LOINC codes all the time, but the investigators do not ask for them, so we do not provide them". This "chicken and egg" situation ultimately leads to non-comparability of test results from various studies, as will be discussed further on. This inevitably leads to errors, as the derivation may well lead and will often lead to a different LOINC code than was used in the laboratory itself.

An interesting point in the above table is that LBMETHOD (method of test) and LBSPEC (specimen type) are "permissible" and until recently had no mandated controlled terminology. CDISC has recently published a list with controlled terms for "METHOD" [13], but it is not limited to lab tests and not synchronized with the list provided by LOINC. Also the list is stated to be "extensible" meaning that every sponsor is allowed to add terms from their own libraries, which of course does not contribute to semantic interoperability. The same applies to the controlled terminology for "specimen type". Although LBCAT (test category) is "expected" in SDTM, no controlled terminology is provided, so every sponsor can categorize tests as they wish, and use their own nomenclature for the categories.

As the lists for LBSPEC and LBMETHOD are very limited and sponsors are allowed to extend them with their own terms, and as LBCAT has no CDISC controlled terminology at all, tests and test results from different studies become incomparable. Imagine the following submitted data from different sponsors for different studies as depicted in table 2

Apparently, the tests are all glucose tests, because LBTESTCD is governed by CDISC controlled terminology with "GLUC" meaning "glucose test". As CDISC does not mandate controlled terminology for LBCAT, and sponsors can add or use their own terms for LBSPEC and LBMETHOD, each row shows different, but somehow similar values for the identifying variables. So the question arises whether these three tests are the same tests or not.

On the basis of the submitted values, it is not possible to say so. Reviewers at the FDA can gain hints by looking at the units (LBORRESU) or the result values (LBORRES) themselves, but this is extremely critical.

However, if in addition the LBLOINC code is given, it can immediately be determined which of these three are identical tests and which are not (table 3).

This example shows that the first and the third tests were the same (2339-0 = "Glucose [Mass/volume] in Blood, Quantitative") whereas the second one was a slightly different test (15074-8 = "Glucose [Moles/volume] in Blood, Quantitative"). When using LBLOINC, essentially the variables LBCAT, LBSPEC and possibly also LBMETHOD become redundant, so that the table can be reduced to table 4

Table 4: Example laboratory test variables from different submissions of different sponsors using only LOINC codes.

Sponsor	LBLOINC
Sponsor 1	2339-0
Sponsor 2	15074-8
Sponsor 3	2339-0

However, a problem that frequently occurs for reviewers at the regulatory authorities is that they do not know

Table 3: Example laboratory test variables from different submissions of different sponsors using LOINC codes.

Sponsor	LBTESTCD	LBCAT	LBSPEC	LBMETHOD	LBLOINC
Sponsor 1	GLUC	CHEMISTRY	BLOOD	QUANT	2339-0
Sponsor 2	GLUC	CHEM	WHOLE BLOOD	ENZYMATIC	15074-8
Sponsor 3	GLUC	CHEMI	BLOOD	HEXOKINASE	2339-0

all LOINC codes by heart and also do currently not have a review system in place where additional information about the LOINC code is automatically generated and displayed. Such an automated lookup of codes will surely be one of the user requirements of a future modern FDA review environment. Our research group has already developed a web service which enables users and systems to look up the meaning of LOINC codes. This web service will be presented in a subsequent paper [14].

Another example comes directly from the CDISC SDTM-IG v.3.1.3 [15]. In the examples for the LB domain (Section 6.3.3.2) we find the following record 6 (table 5).

Table 5: Example record for a lab result from the CDISC-SDTM Implementation Guide.

SDTM Variable	SDTM Variable Value
STUDYID	ABC
DOMAIN	LB
USUBJID	ABC-001-001
LBSEQ	6
LBTESTCD	LYMLE
LBTEST	Lymphocytes
LBCAT	HEMATOLOGY
LBSCAT	DIFFERENTIAL
LBORRES	6.7
LBORRESU	%
...	

Only the variables LBTESTCD and LBTEST have controlled terminology (1:1 relationship), as LBSPEC and LBMETHOD are absent in this case. But exactly which test was meant here? We do not know as in the process of generating the SDTM record, information from the laboratory information system (LIMS) has been lost. If we search for a candidate LOINC code (which was probably used in the LIMS, but lost in the later process) using the RELMA system [7], we will find 162 codes for lymphocytes of which more than 60 are for differential tests (unit fraction, %). Did the test use a manual or automated count with different expectation values? What was the system? Was it blood, body fluid or maybe bone marrow? The SDTM record does not provide this information, which however may be important when evaluating the result, or when comparing values with those from other studies and submissions.

In electronic health record extracts, formatted as HL7-CDA or CEN-13606 or ISO-21090, the usage of LOINC for laboratory test codes is either mandatory or highly recommended. So far, we have not seen any use of CDISC controlled terminology in electronic health records. So when the laboratory information is extracted from such records, the current SDTM-IG requires the code to be mapped

to the CDISC controlled terminology for lab tests which was developed separately without taking LOINC into account. As the above mentioned example shows, this inevitably leads to information loss, also because there is no controlled terminology for LBCAT and LBSCAT, and controlled terminology for LBSPEC and LBMETHOD is limited. In the SDTM record, it is not even visible anymore what the source of the data point was (the electronic health record). Thus it is impossible for the reviewer to find out which test was exactly performed.

A similar problem arises for the units. CDISC has developed its own controlled terminology for units [13]. This list (i.e. not a system) currently contains slightly more than 500 terms, some also being present in UCUM, others being in principal present in UCUM but using a non-UCUM-conform notation, others not being present in UCUM at all. Some even conflict with the UCUM ones. For example, we find the unit "bar" defined as being a "dosing unit" (others are "bag" and "bottle"). In UCUM however, "bar" is a unit for the property "pressure". So in case a pressure measurement was done with the unit "bar" (e.g. a partial oxygen pressure in blood) and stored in an electronic health record, CDISC requires us to translate the "bar" unit into something else (mmHg, torr, Pa, atm, ...) from the CDISC controlled terminology list, as the "original result unit" variable (LBORRESU) is governed by controlled terminology, and "bar" is not in that list as a unit of pressure. This means that even for a good number of these highly standardized and often used UCUM units, the values for the original result (LBORRES) must be recalculated, with the risk of errors. Even worse, "original result" might not be "original" anymore, and there is no way to find out whether this is the case or not, which leads to loss of traceability. It would therefore be better if CDISC recognized UCUM as the base for its controlled terminology for units, possibly extended with very special units not covered by UCUM (and marked as such), which accounts for less than a few percent of all real life cases.

3 Results: The SDTM Laboratory (LB) Domain – An Alternative

3.1 Usage of LOINC Codes instead of LBTESTCD

The analysis of the current SDTM Laboratory (LB) domain shows that the current usage of LBTESTCD (laboratory test code), and the way its controlled terminology

is handled can never lead to comparability of laboratory results from different studies and submissions. In addition, as we have seen, the CDISC controlled terminology is extremely "EHR-unfriendly", as it requires a mapping between test codes and units generally used in EHRs as well as hospital information systems to CDISC codes, which are unfortunately not unambiguous. Due to this fact, these mappings will mostly need to be done manually which inevitably leads to inaccuracies and even errors.

So in cases where the information comes from EHRs and/or hospital information systems anyway, LBTESTCD can better be replaced by LBLOINC (LOINC code). Additionally, the usage of UCUM units for use in LBORRESU (original result unit) and LBSTRESU (standard unit) must be made mandatory. These two measures guarantee that laboratory results of different studies and submissions are comparable.

As the current review systems of the FDA do unfortunately not allow to immediately identify the information belonging to the LOINC code, a "view" can be used temporarily, also displaying the additional information from the LOINC database (i.e. the different components of the LOINC "name") itself, thus leading to a set of new variables which replace the SDTM variables LBCAT (category), LBSCAT (subcategory), and LBSPEC (specimen type) as depicted in table 6

With the following new SDTM variables derived from the LOINC system: LBCOMP (component), LBPROP (property), LBTIMEAS (time aspect), LBSYSTEM (system), LBSCALE (scale) and LBCLASS (class). Each of these variables is governed by controlled terminology, which is the LOINC controlled terminology (not the CDISC one). The following variable values in the table have the following meaning: MCnc = mass concentration, SCnc = substance concentration, Pt = point in time, and Qn = quantitative.

The 5 new variables LBCOMP, LBPROP, LBTIMEAS, LBSYSTEM and LBSCALE are the identifiers for the code, LBCLASS however is a further designator but not an identifier (so not part of the primary key).

A sixth (optional) variable LBMETHOD is not an identifier but a differentiator. The reason is that something different is understood in LOINC under "method" than it is in CDISC. It is the "6th dimension" of the LOINC system, but only used if absolutely necessary to differentiate between two tests that have equal values for the 5 first dimensions. Example values of "method" in LOINC are "agglutination" and "coagulation assay".

This information can easily be generated automatically, as the LOINC organization also provides a database with all the test codes and the corresponding information, and very many LIMS systems have implemented this. Such a database can either be implemented directly in the review software, or be called using a web service. So when a value for LBLOINC is submitted, values for LBCOMP, LBPROP, LBTIMEAS, LBSYSTEM and LBSCALE should not be submitted, as they can be retrieved

from the LOINC database automatically, and be displayed in the tool of the reviewer on request or automatically.

One of the arguments used against the usage of LOINC in CDISC SDTM is that there are always tests that do not have a code. This is correct, but this also the case for the currently used controlled terminology for LBTESTCD (974 terms) which does however not differentiate e.g. between quantitative and qualitative glucose tests. For example, there are at least 36 possible values for "after x hours/days/..." (time aspect) in LOINC, and there is no LOINC code for each combination of the other variables with the time aspect. In such a very seldom case, the table can be similar, but without the LOINC code itself. An example is given in table 7

For the third test, "hexaporphyrin, mass ratio after 12 hours, quantitative measurement in urine", there is no LOINC test code, but due to the controlled terminology for the 5 variables LBCOMP to LBSCALE, the test is uniquely identified, even without the LOINC test code.

3.2 Use of UCUM Units

In the case of electronic health records, or extracts thereof (e.g. in HL7-CDA or HL7-FHIR format), units are usually stored using the UCUM standard notation. So why not submit them as such? The current SDTM-IG forces us to map UCUM units to SDTM units which is first of all not always possible, and foremost time consuming (an automated process is not always possible) and inherently leads to conversion errors for the values. Even worse, the values may become "derived" without any chance for the reviewer to know whether the value is really "original" or whether it has been "derived" (loss of traceability). So in case the source of the data is a laboratory Information Management System (LIMS), or an electronic health record, it definitely makes sense to submit units using the UCUM notation.

A common argument against the use of UCUM units in clinical research is that UCUM does not cover "units" used in preclinical research such as "animals per cage". These proponents of using CDISC controlled terminology for units however mix up the concept of unit with the concept of "annotation". The UCUM specification [8] states the following about this:

"... in chemistry and biomedical sciences, there are traditional habits to write annotations at units or instead of units, such as "%vol.", "RBC", "CFU", "kg(wet tis.)", or "mL(total)". These habits are hard to overcome".

UCUM solves this by using curly brackets for annotations:

"Two alternative responses to this reality exist: either give in to the bad habits and blow up of the code with dimension- and meaningless unit atoms, or canalize this habit so that it does no harm. The Unified Code for Units

Table 6: Example showing the newly proposed SDTM variables for laboratory tests.

Sponsor	LBLOINC	LBCOMP	LBPROP	LBTIMEAS	LBSYSTEM	LBSCALE	LBCLASS
Sponsor 1	2339-0	Glucose	MCnc	Pt	Bld	Qn	CHEM
Sponsor 2	15074-8	Glucose	SCnc	Pt	Bld	Qn	CHEM
Sponsor 3	2339-0	Glucose	MCnc	Pt	Bld	Qn	CHEM

Table 7: Example showing the newly proposed SDTM variables for laboratory tests for the case that there is no LOINC code for a test.

LBLOINC	LBCOMP	LBPROP	LBTIMEAS	LBSYSTEM	LBSCALE	LBCLASS
11217-7	Hexaporphyrin	MRat	24H	Urine	Qn	CHEM
30529-9	Hexaporphyrin	ACnc	Pt	Urine	Ord	CHEM
	Hexaporphyrin	MRat	12H	Urine	Qn	CHEM
50856-4	Hexaporphyrin	SRat	24H	Urine	Qn	CHEM

of Measure canalizes this habit using curly braces”.

Unfortunately, CDISC controlled terminology has given in to the bad habits and is indeed blowing up the list with codes. The UCUM alternative is a much better one. So for example, the unit “g/animal/day” would have the notation “g/animal/d” using curly brackets for the annotations. Similarly, it would enable to use a “unit” like “milligram per bar” using the notation mg/bar where it is clear that “bar” is not the unit of pressure, but a dosing annotation (like a “bar of chocolate”). In order to accomplish this with CDISC controlled terminology, one would need a new term request, with the new term only to become available in the next release of the controlled terminology. As the number of such combinations is almost infinite, this will definitely blow up the code list.

A mapping between CDISC SDTM units and UCUM units was published by the Regenstrief Institute in 2012 [16], using these annotations. This mapping list however is limited to 311 units and seems not to be maintained anymore. Although such a mapping list can be useful, it of course does not make sense to first transform units from electronic health records from UCUM to SDTM, and then later transform them to UCUM again.

So, in our opinion, CDISC should not publish controlled terminology for units, it should publish lists of allowed or recommended annotations to be used in combination with UCUM units.

3.3 The Better Alternative – Example

If we look at the LB domain example in the SDTM Implementation Guide v.3.1.3 (Section 6.3.3.2 on page 137-138), the better alternative using LOINC and UCUM would then be (table 8).

followed by (additional columns – table 9).

as well as by (table 10).

Note that the column “Row” is not part of the SDTM standard, but has been added here for better readability.

In electronic health record extracts, like HL7-CCD, lab tests are almost always identified by their LOINC code. For example, a snippet from an electronic health record in HL7-v3 format is (fig.1).

As such, the information can easily be extracted (e.g. using XSLT) into the SDTM prototype file, allowing automated generation of SDTM records directly from the EHR system.

As more and more information in CRFs comes directly or indirectly from either hospital information systems (where the laboratory test results are also stored using LOINC with units using the UCUM notation), electronic health record systems or electronic health record extracts (e.g. in HL7-v3 format), it is very meaningful to use LOINC and UCUM, either using the LOINC code, or if no code exists for a test (which will be very seldom the case) using the LOINC name (i.e. the combination of the 5 identifiers), to uniquely identify each test. This would also mean an enormous step forward for the FDA, as this creates the opportunity to compare lab values of different studies which is not possible when using the current SDTM and CDISC controlled terminology. It also allows to attach data points from the electronic health record of the subjects to the SDTM record (when stored as XML, such as when using the new CDISC Dataset-XML format [17]), as the former use LOINC coding anyway.

Note that this applies equally to the Vital Signs (VS) domain, as the LOINC system also covers most vital signs measurements (especially those that are currently under controlled terminology by CDISC), and vital signs data in electronic health records typically come with UCUM units. Also here, the advantage would be that FDA reviewers would be able to compare data between studies and submissions which is currently only possible in a very limited way.

4 Conclusions

When hospital information systems or electronic health records are used as the source to provide labo-

Table 8: Example records from the CDISC-IG using the proposed alternative variables (part 1).

Row	STUDYID	DOMAIN	USUBJID	LBSEQ	LBLOINC	LBCOMP	LBPROP	LBTIMEAS
1	ABC	LB	ABC-001-001	1	1751-7	ALBUMIN	MCNC	PT
2	ABC	LB	ABC-001-001	2	6768-6	ALKALINE PHOSPHA-TASE	CCNC	PT
3	ABC	LB	ABC-001-001	3	6768-6	ALKALINE PHOSPHA-TASE	CCNC	PT
4	ABC	LB	ABC-001-001	4	6768-6	ALKALINE PHOSPHA-TASE	CCNC	PT
5	ABC	LB	ABC-001-001	5	26464-8	LEUKO-CYTES	NCNC	PT
6	ABC	LB	ABC-001-001	6	26478-8	LYMPHO-CYTES / 100 LEUKOCYTES	NFR	PT
7	ABC	LB	ABC-001-001	7	26499-4	NEUTRO-PHILS	NCNC	PT

Table 9: Example records from the CDISC-IG using the proposed alternative variables (part 2).

Row	LBSYSTEM	LBSCALE	LBCLASS	LBSCAT	LBORRES	LBORRESU	LBORNRL0	LBORNRHI
1	SER/PLAS	QN	CHEM		30	g/l	35	50
2	SER/PLAS	QN	CHEM		398	[iU]/l	40	160
3	SER/PLAS	QN	CHEM		350	[iU]/l		
4	SER/PLAS	QN	CHEM			[iU]/l		
5	BLD	QN	HEM/BC		5.9	10*9/l	4	11
6	BLD	QN	HEM/BC		6.7	%	25	40
7	BLD	QN	HEM/BC		5.1	10*9/l	2	8

Table 10: Example records from the CDISC-IG using the proposed alternative variables (part 3).

Row	LBSTRESC	LBSTRESN	LBSTRESU	LBSTNRLO	LBSTNRHI	...	LBNRIND	...
1	3.0	3.0	g/dL	3.5	5.0		LOW	
2	398	398	[iU]/L	40	160			
3	350	350	[iU]/L	40	160			
4	374	374		40	160			
5	5.9	5.9	10*3/ μ L	4	11			
6	6.7	6.7	%	25	40		LOW	
7	5.1	5.1	10*9/L	2	8			

ratory test results for clinical research that needs to be submitted to the FDA, the current set of variables for

the CDISC SDTM LB domain fails completely. It is not only necessary to map the LOINC code for the lab test to LBTESTCD and LBTEST (having CDISC controlled

```

<component>
  <observation classCode="OBS" moodCode="EVN">
    <templateId root="2.16.840.1.113883.10.20.1.31"/>
    <!-- Result observation template -->
    <id root="107c2dc0-67a5-11db-bd13-0800200c9a66"/>
    <code code="30313-1" codeSystem="2.16.840.1.113883.6.1" displayName="HGB"/>
    <statusCode code="completed"/>
    <effectiveTime value="200003231430"/>
    <value xsi:type="PQ" value="13.2" unit="g/dl"/>
    <interpretationCode code="N" codeSystem="2.16.840.1.113883.5.83"/>
    <referenceRange>
      <observationRange>
        <text>M 13-18 g/dl; F 12-16 g/dl</text>
      </observationRange>
    </referenceRange>
  </observation>
</component>

```

Figure 1: Example extract from a CCD (Continuity of Care Document) showing a laboratory result coded using LOINC (coding system OID 2.16.840.1.113883.6.1). The lab test code is 30313-1 describing „Hemoglobin [Mass/volume] in Arterial blood“.

terminology), but also to perform a very arbitrary mapping to LBSPEC, LBMETHOD and LBCAT for which either very limited and unfortunately freely extensible CDISC controlled terminology exists or for which there is no controlled terminology at all. As each sponsor will have their own mapping method and vocabulary, comparison of lab test results between different studies and sponsors is thus made nearly impossible. When replacing the latter variables by LBLOINC (and making it at least "expected" in case the source of the data is a hospital information system or electronic health record) together with the new variables LBPROP (property), LBTIMEAS (time aspect), LBSYSTEM (which is more or less equivalent to the current LBSPEC), LBSCALE (scale) and LBCLASS (which essentially corresponds to the current LBCAT, but governed by controlled terminology), each test is uniquely identified, so that it becomes possible to compare laboratory test results between studies and sponsors. Furthermore, the use of UCUM units not only makes comparison between various studies and sponsors possible that have used different units for the same test (UCUM units are easily interconvertible) but also avoids that test values with UCUM units need to be converted to ones with CDISC units. This is not only frequently impossible, but also error prone, and masks whether the value is "as captured" or has been "derived".

As we realize that the current SDTM LB domain cannot be replaced immediately by our newly proposed "EHR-friendly" laboratory domain, we propose that the new domain is named "Laboratory New" with the domain code "LN". Sponsors can then submit laboratory information that was collected in the classic way (i.e. from case report forms) and for which no LOINC or UCUM coding is available using the classic LB domain. They can then submit laboratory data that was received electronically or was retrieved from EHR systems using the newly developed LN (Laboratory New) domain.

References

- [1] ITI Technical Committee IHE, 2011: http://www.ihe.net/Technical_Framework/upload/IHE_ITI_Suppl_RFD_Rev2-2_TI_2011-08-19.pdf
- [2] U.S. Food and Drugs Agency, Study Data Standards Resources, 2014. <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>
- [3] Clinical Data Interchange Standards Consortium. <http://www.cdisc.org>
- [4] See e.g. E.Chappell, "Tracking the evolution of SDTM", 2014. <http://www.formedix.com/2014/02/21/tracking-the-evolution-of-sdtm/>
- [5] LOINC: Logical Observation Identifiers Names and Codes. <http://www.loinc.org>
- [6] Common LOINC Laboratory Observation Codes. <http://loinc.org/usage/obs>
- [7] Regenstrief LOINC Mapping Assistant – RELMA. <http://loinc.org/downloads/reлма>
- [8] UCUM: Unified Code for Units of Measurement. <http://unitsofmeasure.org>
- [9] `ucum_essence.xml`. <http://unitsofmeasure.org/ucum-essence>
- [10] CDA Release 2. Available from HL7: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7
- [11] See e.g. NCI CBIIT ISO 21090 Localization Common Library 2.2, <https://wiki.nci.nih.gov/display/ISO21090/NCI+CBIIT+ISO+21090+Localization+Common+Library+2.2+Release+Notes>
- [12] J.Aerts, Electronic Health Records within ODM: <http://cdisc-end-to-end.blogspot.com/2012/09/electronic-health-records-within-odm.html>
- [13] CDISC Terminology: <http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc>
- [14] J.Aerts, W.Hof, in preparation
- [15] Study Data Tabulation Model Implementation Guide: Human Clinical Trials. Version 3.1.3: <http://www.cdisc.org/sdtm>
- [16] SDTM-UCUM Conversions: <http://unitsofmeasure.org/trac/attachment/ticket/125/SDTM-UCUM%20conversions.xls>
- [17] CDISC Dataset-XML: <http://cdisc.org/dataset-xml>