

Paper DS02

**An open-source project to map lab
terminology and standardize units for
analysis**

Sunil Kumar Pusarla, Omeros Corporation, Seattle, USA

Kevin Viel, Navitas Data Sciences

Dante Di Tommaso, Sanofi, Seattle, USA

Standardize Laboratory Terminology and Convert Units



Collected lab data

with non-standard terms, in original collected units.

Variables

to display for LB_User

original_lbtest original_category original_numeric_result
original_unit

Identify these original data

original_lbtest

original_category

original_numeric_result

Orig. unit

Orig. unit
original_lbtest
original_category
original_numeric_result
original_unit

complete

Search boxes support regular expressions.

LBNAME

LBCAT

LBUNIT

LBPREFU

LBCONV

LB_User

Show 50 entries

Search:

original_lbtest	original_category	original_numeric_result	original_unit
All	All	RANDOM FAKE DATA	All
% Saturation - Iron Studies	blood panel	22.1	volume fraction
% Saturation Iron Studies	blood panel	2.91	v/v
25 Hydroxy Vitamin D	chemistry	26.95	ng%
Abs. Basophils	blood panel	55.03	1000/ul
Abs. Eosinophils	blood panel	29.52	g/l
Abs. Lymphocytes	blood panel	15.53	10^3/ul
Adamts 13 Activity	immuno	2.27	fraction
Adamts13	immuno	1.73	ratio
Adamts-13	immuno	0.23	percent
Adamts-13 Activity Assay	immuno	54.03	L/L
Biocarb	chem	5.11	meq/l
Ca	chem	37.47	meq/l
Ca Oxy Crystals	urine	0.35	/lpf
Haemoglobin	hema	24.8	g/100ml
Hemaglobin	hema	0.07	g/l
Hemoglobin Monomer (Subunit)	hema	52.84	mu/l
Mch - Mean Corpuscular Hemoglobin	hema	4.48	mg%

Introduction

- Conversion of lab data from original to standard units is a common task in clinical trials
- The task involves translation of at least three variables (fields)
 - Lab analyte name (LBTEST)
 - Lab category (LBCAT, a surrogate for specimen [LBSPEC])
 - Lab unit (LBORRESU)
- Various resources exist, but few freely available sources are conducive to programmatic approaches

Issues and constraints

- Time and resource constraints
- Voluminous non-standard laboratory data essential to regulatory applications
- Varying international terminology and measurement conventions
- Study-specific requirements
- Mandates for unimpeachable quality and transparency of data processing and analysis.

Main Challenges

1. Map novel terminology to standard terminology, primarily for test names and original result units.
2. Accumulate a database of conversion factors.
3. Deliver within a defined timeline and budget.

Example: Regulatory requirements

FDA data validation rule **FDAB030** (rule ID SD0007):

“Standard Units (--STRESU) must be consistent for all records with the same Short Name of Measurement, Test or Examination (--TESTCD), Category (--CAT), Subcategory (--SCAT), Specimen Type (--SPEC) and Method of Test or Examination (--METHOD).”

FDA, *The Validator Rules v1.5*, March 2021, <https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources>

Sources of Information: Controlled Terminology

- National Cancer Institute (NCI) provides preferred terminology (LBTEST, LBTESTCD, LBORRESU) and CDISC synonyms

Sources of Information: Conversion between units

- <http://unitslab.com/>
- <https://dwjay.tripod.com/conversion.html>
- The NEJM series based on the case records of Massachusetts General Hospital

Laposata, Michael, SI Unit Conversion Guide, NEJM Books, Boston, MA, USA. 1992.

Kratz, Alexander; Lewandrowski, Kent, Normal Reference Laboratory Values, New England Journal of Medicine, v. 339, n. 15, October 8, 1998, pp. 1063-1072.

Kratz, Alexander; Ferraro, Maryjane; Sluss, Patrick; Lewandrowski, Kent, Laboratory Reference Values, New England Journal of Medicine, v. 351, n. 15, October 7, 2004, pp. 1548-1563.

Why Open Source?

- These sources are fragmented and not available in a format that reflects FAIR data principles

FAIR data, Wikipedia, https://en.wikipedia.org/wiki/FAIR_data.

- Errors and variability in the standardization of terminology and conversion of test results can be difficult to detect and pernicious
- Should choosing between a conversion factor of 40 or 40.1 or 40.078 compromise patient safety or shift a public health decision?

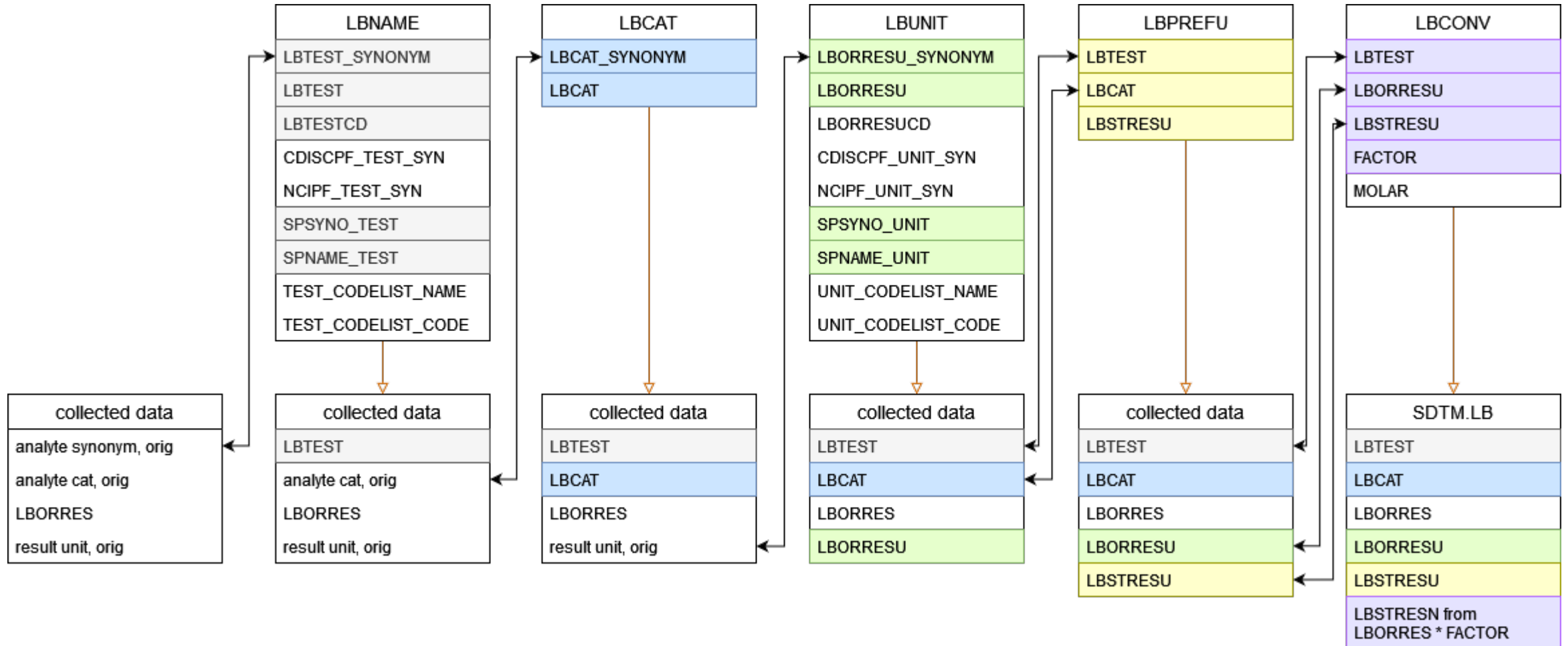
Why Open Source? (continued)

- Implementation of varying approaches that arrive at the same values is (independent) Validation.
- Improved robustness
- Improved reliability
- Improved transparency of existing features
- Diverse subject-matter expertise

Why Participate in this Open Source Project?

- Medicine and public health depend on high quality data and analyses.
- Wide variety of contributions that you can make
- Share and learn technical skills
- Increase subject-matter expertise
- Build your reputation within your professional community

Workflow



Hierarchical approach

- Enterprise standards.
 - Project configuration or conventions.
 - Study configuration or conventions.
-
- Only Enterprise data are required.

Study-level data overrides Project-level data, which overrides Enterprise-level data

Example: Enterprise-level synonyms

The following spellings are both not uncommon, but clearly are intended to be HEMOGLOBIN:

HAEMOGLOBIN

HEMAGLOBIN

Example: Study-level synonyms

- Misspelling and (curious) alternatives occur in the data
- Rare (one off's), i.e. data from one local lab in one trial
- Not, yet, established enough to be Enterprise-wide.

Actual examples:

Billirubin -> Bilirubin

% retic -> Reticulocytes/Erythrocytes

Synonyms to Controlled Terminology

STEP 1: TEST NAME, LBNAME TABLE

cholesterol -> Cholesterol

STEP 2: TEST CATEGORY, LBCAT TABLE

Chemistry -> CHEMISTRY

STEP 3: RESULT UNITS, LBUNIT TABLE

MG/DL -> mg/dL

Preferred unit and conversion factor

STEP 4: PREFERRED STANDARDIZED UNITS, LBPREFU TABLE

Cholesterol, CHEMISTRY, mmol/L

STEP 5: CONVERSION FACTORS, LBCONV TABLE

Cholesterol 1 mg/dL = 0.02586318 mmol/L

LBORRES = "180" LBORRESU = "mg/dL" -> LBSTRESN = 4.6554
LBSTRESU = "mmol/L"

What is the correct number of significant digits?

Analyte-invariant versus Analyte-variant conversions

- Conversions that do not depend on the identity of the analyte are “analyte-invariant”.
 - Original and preferred units match
 - Original and preferred units are the same category, mg/dL to g/L.
- Conversions that depend on the identity of the analyte are “analyte-variant”.
 - Mass (grams) converted to Molecular mass (moles)

Analyte (in)dependence

Search boxes support regular expressions.

LBNAME

LBCAT

LBUNIT

LBPREFU

LBCONV

LB_User

Show 50 entries

Search:

lbtest	lborresu	lbstresu	factor	molar
^(bili bicar calc)	^mmol	All	All	All
	mmol/L	mmol/L	1	0
	mmol/L	umol/L	1000	0
Bicarbonate	mmol/L	mEq/L	1	1
Bicarbonate	mmol/L	mg%	6.10168	1
Bicarbonate	mmol/L	mg/100mL	6.10168	1
Bicarbonate	mmol/L	mg/L	61.0168	1
Bicarbonate	mmol/L	mg/dL	6.10168	1
Bicarbonate	mmol/L	ug/mL	61.0168	1
Bilirubin	mmol/L	mg%	58.467	1
Bilirubin	mmol/L	mg/100mL	58.467	1

Implication of LBCAT

- The conversion will be mathematically valid if the “wrong” preferred unit is used, but the scale might be “off”.
- For instance, conventionally, an analyte might be measured in the ng range in blood, but in the pg range in urine.
- Multiple LBPREFU entries with distinguishing LBCAT values enable valid conversions suitable for lab categories

Upcoming tasks

- SDTM Terminology is versioned and released quarterly.
 - Standardize Labs should support versioned CDISC terminology, as well as versioned sponsor extensions.
 - The application should reconcile any conflicts between sponsor extensions with SDTM Terminology.
- Incorporate LOINC.
- Standardize terminology for qualitative results, such as for urinalysis results.
- Confirm and protect expected behavior of the application with robust, automated tests.

QUESTIONS or FEEDBACK?

Thank you!

Sunil Kumar Pusarla, Omeros Corporation, Seattle, USA

Kevin Viel, Navitas Data Sciences

Dante Di Tommaso, Sanofi, Seattle, USA