ELM Cumul Project position papers

2. Why CUMUL?

Jack Bierens de Haan

Summary

LOINC®, an English language product, cannot in any useful way be translated into another language as it intricately mixes code and information. Neither can RELMA®, the LOINC mapping assistant utility, be adapted to operate on other than English local files. The "native" form of LOINC therefore cannot be expected to be widely adopted in other than English speaking environments for the mapping of local test master files. To overcome this limitation, sets of index names were developed in several languages under the name CUMUL. These names point to LOINC numbers and allow bypassing the LOINC name in most cases. Selected units of measurements were added to help keep names short, subclass tags were nested under LOINC classes and the syntax was adapted such that CUMUL names can also be used for patient test reporting.

Introduction

This is the second paper in the ELM Cumul project papers series[1] and aims at answering the question: "if LOINC[2] is the only currently available candidate for a universal test directory then why not just recommend that laboratories implement LOINC numbers in their local test master files and move on to more thrilling communication issues?"

LOINC is written in English

The reason is that such a bare recommendation would meet with little success in other than English-speaking environments, that is in most European countries. Implementing LOINC numbers involves finding an equivalent fully specified LOINC name to each local test name and assigning the LOINC number to that test, an operation which is re-

ferred to as *mapping*. The exercise assumes a good knowledge of LOINC syntax and semantics and an ability to read "between lines" of the local file (that is equating apparently different terms and making up for missing terms, in other words *interpreting*, based on expert knowledge) but it is first of all based upon just reading the lines and, in any case, does not assume any multilingual skill.

Even so, mapping is not a straightforward task because a directory with exhaustive pretension will tend to list all existing tests regardless of their relative relevance or frequency and will therefore be ten to twenty times as large as the average local file, often leaving the user hesitating between two like entries or wondering whether he should submit a new one. The issue was recognised from the outset by the promoters of LOINC and tentatively addressed by their provision of RELMA[3], a browsing/mapping program which will take the user by the hand and lead him stepwise towards mutual recognition between LOINC and local jargon, enabling him to assign LOINC numbers to his local tests in a tidy way.

Because the 29'000 LOINC entries include 18'300 distinct words in a 368'200 English word index, officially recommending LOINC is likely to remain a verbal exercise if nothing is done to overcome the language barrier where this exists.

Translating LOINC ?

As explained previously (see part 1) LOINC is a multi-column table exclusively populated with coded variables where code stands for information (e.g. SER stands for Serum). From this point, the LOINC table will be called a *database* to distinguish it from other tables. Indeed, individual codes paired with their corresponding information entity are listed in two-column coding tables (e.g. for Sytem, Scale. Method, etc.). The information in the coding tables can be translated but not the code. Translating a code is pointless, it amounts to re-coding, that is ignoring the very standard. Translating the information in the coding tables, on the other hand, can be an option, but of limited interest in this case because the readability of the LOINC database owes much to the information being *evoked* by the mnemonic code, such that in practice the user seldom resorts to the coding tables.

In this respect the *Component* terms in the first part (and

---

[1] see ELM cumul project position papers 1. LOINC vs. NPU

[2] Logical Observation Identifier Names and Codes

[3] Regenstrief LOINC Mapping Assistant

subparts) of the fully specified name pose a special problem because there is no table for them. They serve as both code and information. If these are to be translated, the translations will need to be acccomodated into a newly appended *information* column with the english term being retained in a *code* column. Then, if the coincidence of code and information is lost and the reader's language is remote from English, the LOINC table will look unfamiliar and present many traps.

In other words, a translated LOINC alone would require the user to continually refer to the coding tables i.e. the product would not be any friendlier than the native LOINC.

Translating RELMA ?

The RELMA utility provides a rather complex, multifunctional Windows-based interface in line with two source databases,

- LMOF[4] which contains relevant information imported from the local laboratory test file, and

- LOINC.

The interface essentially allows to pull records from either side with a view to relate each of the LMOF entry to an appropriate LOINC entry, using elaborate search-filter-select features.

The package includes a utility which is to be applied prior to mapping and intended to spot in LMOF those terms which are not part of the LOINC vocabulary, and to allow either replacing them with LOINC compatible terms or removing them. This pre-mapping utility is intended mainly for getting rid of locally coined abbreviations and acronyms.

Could RELMA be translated and applied to an LMOF written in another than English language? The interface mask and the User's Manual of RELMA could obviously be translated. As already mentioned, translated coding tables could be prepared for each part and subpart of the LOINC names, including the Component terms.

This being done, one could think of adapting the pre-mapping utility to scan for anything that does not belong to this *translated* LOINC vocabulary, and the RELMA engine to look for *translated* information when comparing LMOF terms with LOINC terms. But much of RELMA's ergonomy rests on the ambiguity which LOINC can maintain between component codes and names in English. The character string that RELMA brings forward for the user to make a decision remains a concatenation of LOINC codes.

In other words, translating and adapting RELMA would be a considerable effort with little outcome.

Adding a layer to LOINC

The fact that it cannot in any useful way be translated into another language was soon perceived as a major obstacle to the promotion of LOINC in Switzerland, a country with four national languages. To the the CUMUL[5] group, self-entrusted with this promotion, the only way out seemed to try and provide LOINC records with some sort of language-specific tags that would allow indexing the table for other than English speaking users.

This amounted to naming tests once again, an issue which very much looked like re-inventing the wheel (see part 1) . The scope and conditions of the task however made this considerably simpler:

- the entities to be renamed were 6-part LOINC entries rather than real world tests;

- the new names were to appear *on the top of* the LOINC directory, leaving the original record available for scrutiny (by any user with sufficient basic English notions);

Then, to meet the indexing purpose in a mapping context, the tags had to be:

- unique (though not necessarily fully specified);

- as short as possible, for ready identification and manipulation

- highly readable, i.e. they had to "look familiar" to the person in charge

The idea was to build in each language an index of names that could be directly related to LOINC numbers yet, for most purposes, bypassing fully specified LOINC names. Hence the concept of *layering* a set of multilingual indexes on the top of LOINC to facilitate access to and use of the database.

The reportable name issue

So far Cumul efforts had aimed at opening up the LOINC-mapping of local test files to other than English speaking users. The objective of mapping, it must be reminded, is to allow computers to recognise the same test behind names which are spelled differently, not to impose a standard spelling.

It was soon realised however that, with the conditions set as above for terseness and legibility, the newly coined names became close to a longed for species: the *standard reportable name*. The reportable name of a test is that by which it is most often referred to, as e.g. on a patient test report or in conversation. Lack of standards in this respect is the very source of the need for such directories as LOINC. And the structure of LOINC names sufficiently demonstrates the difficulty to fully identify tests in text. On the other hand, the average patient test report is based on names which do not exceed 30-40 characters yet seem good enough for the target user (if this is a human). Now "good enough" is maybe not optimal. Physicians in general and a majority of labo-

---

[4] Local Master Observation File

[5] CUMUL is not an abbreviation; it is the name of a medical informatics project initiated in 1997 in Geneva (Switzerland); see http://www.cumul.ch

ratories would welcome more homogeneity, i.e. some sort of guideline for naming tests.

Although the working group did not feel entrusted with any standard-setting mission it decided to seize the opportunity and propose a test naming system that could potentially be used for both mapping and reporting purposes.

The unit issue

One reason why shorter names are acceptable in routine work is because the *unit* will often tell about the *kind of property*, the *scale* and the *time aspect*. The unit is a highly informative item. Because the unit is not part of the LOINC name, it was decided to add this to the CUMUL layer and drop any redundant information from the CUMUL name to help keep it as short as possible.

Admittedly, assigning a unit to a LOINC entry can raise problems as it will in fact restrict its use and deprive the same test expressed in alternative units from a deserved LOINC reference. For example, assigning the unit *g/L* to a component with the property *mass concentration* would leave a test expressed in *mg/100mL* without CUMUL reference or, stated otherwise, if a *substance concentration* result can be expressed either as *0,00x mmol/L* or as *00x nmol/L*, CUMUL will impose one of the two alternatives, leaving the other one without a CUMUL reference.

This illustrates perhaps the reason why LOINC dropped the unit altogether but it also points to a weakness of LOINC which can deliver correctly *identifiable tests* yet stops short of delivering readily *comparable results*.

Considering the fairly homogeneous European context, the Cumul working group decided to take the risk and post the unit, thereby departing from the more neutral, fully pragmatic attitude of LOINC.

The abbreviation issue

Abbreviations in the form of capitalised initials of complex terms are extremely common in laboratory medicine. The policy of LOINC is very restrictive about their use for component names and rightfully so because this would defeat the readability feature of the database. Abbreviations would require a specific coding table. CUMUL followed the same policy for the same reason and also because the set of common abbreviations in other than English languages will most often be a mix of national and English acronyms, the exact significance of which is lost, making the confusion even worse.

The reporting issue

The above decision to develop short test names that could be used for both mapping and reporting deeply changed the scope of the Cumul project. It was then realised that, with only a few more steps, the new reportable test names could be made to auto-arrange into a meaningful report without the need of a specific sort routine. With such a

feature in the Cumul layer, any string of LOINC-numbered patient test results would become autoreportable, i.e. Cumul would provide the appropriate ordering sequence and hierarchical structure to produce an acceptable layout.

The required steps were based on following assumptions:

- Two levels of test-grouping, e.g. *chapter* and *subchapter*, are sufficient for an acceptable report layout, the subchapter level being that below which the test order becomes arbitrary.

- Within the above hierarchy, some tests still need to presented as ordered subsets (e.g. electrophoresis fractions, erythrocyte indices, etc.).

The first feature was achieved by carrying over and building upon the LOINC *class*. The second feature was achieved by setting appropriate rules for the name construction.

The class/subclass issue

LOINC distributes tests into *classes*, or conventional chapters, to narrow down the portion of the database which the user must browse when looking for a specific test. This is not an index key field because tests can be uniquely identified without resorting to the class. Besides, the class, a concept which more or less overlaps that of scientific discipline (from an undefined mix of analytical, biochemical and clinical points of views), is not sufficiently objective to allow for a consensus definition and for undisputed assignation of individual tests.

On the other hand patient test reports are always arranged into "chapters" following a similar if not identical rationale (e.g. most reporting schemes will combine BC, HEM and CELLMARK classes under "Haematology"). Using LOINC classes to organize the patient test report into meaningful subsets may thus look awkward but definitely not absurd provided that the order of such subsets is not random.

Cumul therefore endorses the LOINC class but sets the sequence in which these are to be reported.

While some LOINC classes would better be combined for reporting, others are definitely too broad and need further subdivision. Hence the introduction of a new Cumul *subclass* tag for some of the classes. The subclass is an arbitrary group whithin which the order of tests layout is no longer significant, i.e. it can just as well be alphabetical. Subclasses were developed for five classes (see tables 2-6). There is not necessarily a subclass for each test in a class, i.e. some tests may remain not subclassified.

Cumul syntax

The syntax must be such that tests in a report message, when sorted by *name*, within *subclass* and *class*, and with the *unit* posted, will become laid out as a readily understandable, logically structured, meaningful report. Names should be as short as possible, yet with scarce use of abbreviations.

The approach is freely adapted from C-NPU principles, i.e. departing only from these to avoid:

- dispersing tests that ought to remain grouped;

- producing redundancy (e.g. with the *unit*, when this is posted);

- expressing self-evidence or a wide consensus.

As the above criteria are rather subjective, and entail much simplification, the resulting name cannot be as specific as the *fully specified LOINC name* which therefore remains the ultimate reference for file mapping in case of doubt or ambiguity.

The CUMUL name follows the format:

System(spec)–**Component**(spec); "**Property**"; **Method**

where (spec) stands for (specification) and Property is purposely written between quotes; the separators semicolon (;) and comma (,) are followed by a space, for improved readability.

The terms System, Component, Property and Method are further detailed below.

System

The term System complies with the following rules:

- The primary specimen comes first, optionally followed by a bracketed subsystem as in *U(Sed)*, *Bld(Wbc)*, etc.;

- For excretions, the patient is not considered a system: such notations as *Pt(U)*– or *Pt(F)*– are substituted for *U*– and *F*– respectively, as the *unit* will tell when this is an excretion (g/d, umol/d, etc.);

- Frequent values are abbreviated; the choice of abbreviations was oriented by NPU[6] rather than LOINC policy abbreviations will be taken over, as in Table 1 below (yet with only the initial capitalized).

- The value *SER/PLAS* is substituted for *S*.

Table 1 - Abbreviations for systems

Amf	amniotic fluid	Mec	meconium
Asp	aspirate	Pt	patient
Bar	arterial blood	Pcf	pericardial fluid
Bca	capillary blood	Ptf	peritoneal fluid
Bve	venous blood	P	plasma
Bro	bronchial aspirate	Plf	pleural fluid
Csf	cerebr. spin. fluid	Sal	saliva
Cvm	cervical mucus	Sep	seminal plasma
Diaf	dialysate	S	serum
Duf	duodenal fluid	Spt	sputum
Rbc	erythrocytes	F	stool
Gas	gas	Sw	sweat
Gast	gastric contents	Trc	thrombocytes
Lkc	leukocytes	U	urine
Lyc	lymphocytes	Vom	vomitus
Mar	marrow (bone)	B	whole blood

Component

The term Component complies with the following rules:

- When comprising several keywords these will be arranged in order of decreasing "weight", separated by commas (,) unless this is the natural order of the language (e.g. *S—Amylase, pancreatic...*);

- New keywords may be added ahead of the name to ensure a structured layout and the integrity of grouped results sets (as, e.g. in : *Bld-erythrocytes, indices, mean volume ...*);

- Terseness takes priority over systematics as long as the meaning remains clear (e.g. the term *virus* is omitted when this generates no ambiguity);

- A small number of widely acknowledged abbreviations are authorised; in English and in order to avoid duplication of work, LOINC abbreviations will be taken over, as in table 2 below.

Table 2 – Abbreviations for component parts

ADP	adenosine diphosphate
ATP	adenosine triphosphate
DNS	deoxyribonucleic acid
ENA	extractable nuclear antigen
HDL	high density lipoprotein
HIV	human immunodeficieny virus
HLA	histocompatibility leucocyte antigen
HTLV	human lymphotropic virus
Ig	immunoglobulin...
IGF	insulin-like growth factor
INR	international normalised ratio
LDL	low density lipoprotein
RNP	ribonucleoprotein
RNS	ribonucleic acid
TRH	thyrotropin releasing hormone
VLDL	very low density lipoprotein

Property

The term "Property" calls for two remarks:

- It is set beween quotes because the term is taken in a wider and more flexible acceptation than the standard term, more particularly to include the reference component when the test component is expressed as a fraction or a ratio (e.g. *hemoglobin fraction*; *ratio to total CK*; *ratio to creatinine*; etc.).

- The term is posted optionally, i.e. only when there is no European consensus and it is not implied by the *unit*.

Method

The term Method calls for two remarks:

- It is posted only when deemed useful for clinical interpretation (LOINC rule) or to distinguish results likely

---

[6] NPU in this context stands for the project no710/18/87 of the IUPAC/IFCC Commisssion on Nomenclature Properties and Units

to be obtained by different method for the same component in the same report (as sometimes occurs in infectious serology)

- A small number of widely acknowledged abbreviations are authorised; in English and in order to avoid duplication of work, LOINC abbreviations will be taken over as in Table 3 below.

Tabl 3 - Abbreviations for methods

AGGL	agglutination
CFR	complement fixation reaction
CIE	counter-immunoelectrophoresis
COAG	coagulometric
CZE	capillary zone electrophoresis
ECL	electrochemiluminescence
EIA	enzyme immunoassay
ELP	electrophoresis
ENZ	enzymatic
FPIA	fluorescence polarisation immunoassay
GC	gas phase chromatography
GCMS	gas chromatography mass spectrometry
GEN	genomics w/wo sugnal/target amplification
HA	hemagglutination
HAI	hemagglutination inhibition
HPLC	high performance liquid chromatography
IB	immunoblot
IE	immunoelectrophoresis
IEF	immunoelectrofocusing
IF	immunofluorescence
IFX	immunofixation
IMM	immunological
IR	infrared spectrometry
ISAGA	immunosorbent agglutination assay
LATX	latex microparticle agglutination
MSCOP	optical microscopy
RIA	radioimmunoassay
RIBA	recombinant immunobinding assay
RID	radial immunodiffusion
STICK	dipstick
TLC	thin layer chromatography

Chapter and subchapter

The LOINC Class, a mandatory field (field-9), is retained as a major sort criterion for the presentation of chapters. LOINC values are ordered in a conventional sequence as in table 4 below.

The CUMUL Subclass is optionally posted. It can take the values, ordered in a conventional sequence within each class, as in Tables 4.

Unit

LOINC Field-29 **ipcc_units** is meant to receive in the future a C-NPU recommended unit. Until this occurs, CUMUL provides a field that may or may not be distinguished from the latter and holding:

- A consensus unit which will mostly match IUPAC-IFCC recommendations, yet sometimes depart from strict compliance, as e.g. mass fraction or number fraction ex-

Table 4 - main LOINC classes

HEM	Haematology
BC	Cell counts
CELLMARK	Cell markers
COAG	Haemostasis
BLDBK	Immunohaematology
CHEM	Chemistry
CHAL	Function tests
SERO	Autoimmune serology
ALLERGY	Allergens
CHALSKIN	Skin tests
FERT	Fertility
DRUGDOSE	Drug dose
DRUG	Therapeutic drug monitoring
TOX	Toxicology
UA	Urinalysis
MICRO	Microbiology
ABXBACT	Antibiotic susceptibility
CYTO	Cytology
PATH	Pathology
SURGPATH	Histopathology
HLA	Tissue typing
MOLPATH	Genetic anomaly

Table 5 - CUMUL subclasses

Subclasses for Cell Counts	
difa	Automated leucocyte differentiation
difm	Manual leucocyte differentiation
Subclasses for Chemistry	
elec	Electrolytes and blood gases
ppt	Precipitates
metab	Metabolites
enzy	Enzymes
enzel	Isoenzyme electrophoresis
prot	Proteins
protel	Protein electrophoresis
protim	Protein immunoelectrophor., immunofixat.
lipid	Lipid metabolic markers
hematin	Hematinics
Subclasses for Allergy	
pmic	Pneumallergens: microorganisms
pdus	Pneumallergens: acarians, dusts
pgra	Pneumallergens: grass pollens
pwee	Pneumallergens: weed pollens
ptre	Pneumallergens: tree pollens
ppsc	Pneumallergens: pollen screens
pani	Pneumallergens: animal residues
pveg	Pneumallergens: vegetal residues
ppro	Pneumallergens: occupational exposure
tveg	Trophallergens: vegetal
tani	Trophallergens: animal
iven	Inoculated/absorbed allergens: toxics
ipar	Inoculated/absorbed allergens: parasites
idru	Inoculated/absorbed allergens: drugs
imisc	Inoculated/absorbed allergens: miscellan.
Subclasses for Microbiology	
sbac	Bacterial circulating antibodies/antigens
smyc	Fungal circulating antibodies/antigens
spar	Parasitic circulating antibodies/antigens
svir	Viral circulating antibodies/antigens
direct	Direct determination of infectious agent
cult	Culture
Subclass for Urinalysis	
strip	Dipstick chemical testing

pressed in **%**; litre noted **L** to avoid confusing with the number one (1); number concentration expressed without entity symbol, as the nth power of 10, noted **10En/L**; the letter **u** to symbolise the greek letter m not always available on current character sets.

- In most cases the unit expresses the measured property which allows omitting the latter.

- Where no unit is present the result may be a serological titre **[titre]**, another dimensionless value **[no dim]** or it may be posted on a discrete **[ord]** or nominal **[nom]** scale.

It may not be sound practice to use the same field to express such different LOINC concepts as the *unit* (which is a reference property with the value 1), the *property* (for the serological titre) and the *scale.* But it is just a fact that, in the above case, the three concepts are mutually exclusive and the resulting expression is unambiguous and concise.

Discussion

An index of CUMUL names in the user's language is an invaluable productivity tool for anyone engaged in the LOINC-mapping of local test master files.

CUMUL names are meant to look familiar to the average user of laboratory tests, a feature that was deliberately relegated by LOINC designers and NPU experts as well, both insisting that fully specified names are *not* meant for the end-user and arguing that this is the way it should be.

The designers of CUMUL, on the contrary, realised that the conditions set for test mapping could be very close to those set for test reporting. They borrowed from NPU nomenclature until the test name became obscured by standard syntax and semantics, after which they made choices that may look arbitrary but were always guided by the requirement of readability.