

# Guidelines for modelling Units of measure in SDMX

---

Version 1.0.draft

SWG - Unit of measure task group

October 2024

## 1. Introduction

---

The importance of unit of measure in statistical data modeling is to assure consistency and accuracy of the data. Measurement units are used to quantify data and to offer an immediate mechanism to assess the comparability and computation scope of the data.

When comparing statistical data indicators, accurate and consistent unit of measure conversions are crucial for data integrity and meaningful analysis. This is why guidelines for Unit of Measure metadata modeling are essential. These guidelines provide a framework for organising and standardising the way units of measure are represented and managed in data models.

Also, modeling unit of measure and its associated statistical metadata are important for the exchange of data and metadata between organisations. As part of the Statistical and Metadata Exchange Standard (SDMX) consortium, tailored guidelines are developed that recommend harmonised cross-domain concepts and terminology to improve the efficiency of data exchange.

This content-oriented guideline for Units of measures is destined to statistical data modelers working with multidimensional data, primarily in SDMX – but with a scope that can go beyond (e.g. datawarehousing in general). The patterns proposed in this paper are anchored in existing and well-established international standards in the fields of science and engineering.

Beyond the well established unit of measure standards, there are two novelties addressed in this guideline: on one hand it describes dimensions and their corresponding *units of measures* that are typical to economics and social statistics, and on the other hand it deals with data structured into multi-dimensional cubes.

The measurements particular to the fields of economics and social sciences are *monetary value* and numerosity of various sets (e.g. population, enterprises). The main aim is to keep the spirit of the existing standards while integrating the new dimensions, that is facilitate comparability of the data across a wide spectrum of data sources and help identify the computation scopes in which data can appear.

The second novel element of the guideline is introducing the perspective of a multi-dimensional cube, living in a data-warehouse of multiple, interrelated cubes, and in the ultimate idealized scenario as part of a universe of connected data warehouses. The main standards that traditionally govern the rules for units of measures have a simple model to represent measured quantities  $Measure = \{x\}[unit\ of\ measure]$ . In this guideline we discuss cases when both the *measure* quantity and the *unit of measure* emerges as a combination of dimensions and attributes (jointly referred to as components). In practice, the difficulty is indeed in delineating components that are part of the measure and components that describe the units of measure, and building a consistent set of such components that can be recomposed systematically into the compact representation of the existing standards. The guideline presents cases and provides recommendations for various concepts that impact the measurement or the allocation of the units of measure (e.g. unit multipliers, transformations and adjustments, scaling facts and deflator types) - including code-lists or code-list templates.

By following these guidelines, organisations will be better equipped to ensure data consistency, improve interoperability, and enhance the overall quality of their data.

## 2. Alignment to existing international standards for Units of Measure

---

There are recognized standards for representation for units of measure. The International System of Units (SI), International System of Quantities (ISQ), and Unified Code for Units of Measure (UCUM) are the most frequently used and acknowledged standards. We align the Unit of Measure guidelines with the current standards as we create cross-domain code-sets to represent units of measurement for the statistical community.

## ■ **Système International (SI) for quantities and units (ISO 80000)**

ISO 80000 is an international standard describing the International System of Quantities (ISQ) and the International System of Units (SI). ISO 80000 is a comprehensive and updated series of standards covering various fields of science and technology. ISO 80000 also incorporates some amendments that were made to ISO 31 and ISO 1000 in the past. It also defines the symbols and rules for writing quantities and units in a consistent and coherent way. It is developed and published jointly by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

## ■ **The Unified Code for Units of Measure (UCUM)**

The Unified Code for Units of Measure (UCUM) is a code system that aims to include all units of measure that are currently used in various fields of science and engineering. It is designed to enable clear and accurate electronic communication of quantities and their units, especially in terms of driving interoperability. It can help avoid confusion between different systems of measurement, such as metric and imperial, or different conventions for writing units, such as using dots or slashes to indicate different parts of a division. UCUM has been adopted by many international organizations and standards, such as IEEE, DICOM, LOINC, and HL7, and is also part of the ISO 11240:2012 standard.

## 2.1. Using dimensional analysis

---

In engineering and science - as a companion of the mentioned standards - dimensional analysis is systematically used to determine units of measures of measured and derived quantities. Dimensional analysis comprises of the analysis of measured quantities to identify their base dimensions (such as length, mass, and time) and to associate a unit of measure (such as miles or kilometres) with the quantity such that the measured quantity is expressed as multiples of the unit of measure. Additionally, dimensional analysis sets out the rules to working with measured quantities in calculations and comparisons.

Dimensional analysis is a convenient tool to build a rigorous system of measurement, and serve as a validation tool for calculations and comparisons. The main rules of dimensional analysis/calculus can be summarised as follows:

- only measured quantities sharing the same units of measure can be directly compared, or added to (subtracted from) each other
- measured quantities sharing the same base dimension are said to be *commensurable*; their roles as measured quantity vs. unit of measure are interchangeable (e.g.  $in = 2.54cm \Rightarrow cm = 1/2.54in$ )
- multiplication and division is possible between quantities of different units of measure, with the resulting quantity carrying the unit of measure that results from the same operations carried on the units of measure (using the conventional rules of exponentiation:  $\frac{m/s}{s} = m/s^2$ )
- derivation/differentiation of a quantity over another corresponds to a division, whereas integration corresponds to a multiplication in the units of measure

Just like in engineering and science, statistical databases present measured quantities (and estimates and derivations involving those measured quantities), so it is appealing to consider the usage of dimensional analysis to improve the analytical utility of the presented data. The analytical utility would come from the usual properties of dimensional analysis: if the data producer associates units of measure in line with dimensional analysis rules, then they can convey useful immediate information of the scopes in which the data can be used in, of limits of comparability, and can drive deeper insights when using complex transformations of the data. The systematic representation of units of measure in data models can avoid clumsy reference metadata and methodology lookups.

There are two notable points where the existing standards allow for some interpretational flexibility: treating ratios of commensurable quantities (i.e. quantities that share the same base dimension) and handling the numerosity of sets. We will discuss these two cases in the following sections.

## 2.2. Deriving units of measure with dimensional analysis

---

The examples below demonstrate how dimensional analysis works, and discusses two distinct scenarios of data derivation: ratios of commensurable quantities and ratios of non-commensurable quantities.

In the following example we use three measured quantities of an imaginary country:

- population: 100 *persons*
- gross debt: 5000 USD

- reserves: 4000 USD

Two of these quantities represent *monetary value* (with USD as the associated measurement standard) and one represents *numerosity* (with person as the measurement standard)

Using rules described by dimensional analysis we can derive new measures, by computing ratios of non-commensurable quantities :

$$\text{Debt to population ratio} = \frac{\text{gross debt}}{\text{population}} = \frac{5000 \text{ USD}}{100 \text{ person}} = 50 \text{ USD/person} \quad (1)$$

$$\text{Reserves to population ratio} = \frac{\text{reserves}}{\text{population}} = \frac{4000 \text{ USD}}{100 \text{ person}} = 40 \text{ USD/person} \quad (2)$$

These are all straightforward cases, following from the 3rd bullet-point on dimensional analysis, mentioned above. The measured quantities are 'Debt to population ratio' and 'Reserves to population ratio' respectively, and in both cases the associated unit of measure is USD/person.

However in the case of a 'Debt to reserves ratio', i.e. a scenario in which we calculate ratios of quantities that share the same unit of measure (and hence are commensurable) we have two competing representation options (for the same numerical figure<sup>[^1]</sup>).

### 'Unitless number' approach:

$$\text{Debt to reserves ratio} = \frac{\text{gross debt}}{\text{reserves}} = \frac{5000 \text{ USD}}{4000 \text{ USD}} = 1.25 \text{ USD}^0 \quad (3)$$

In this scenario the value 1.25 is stored for the measure 'Debt to reserves ratio' and it is unitless in the sense that the original units cancel out if using the familiar arithmetics  $\text{USD}/\text{USD} = \text{USD}^0 \stackrel{\text{def}}{=} 1$ . The *def* qualifier highlights that it is by convention that we consider such quantities relieved from the burden of being tied to their original dimension and are treated as pure, dimensionless numbers.

The struggles and workarounds statistical producers take to fill the void created by the 'disappearing' unit of measure in the 'unitless number' approach is telling. Often the component for the unit of measure in the database is filled by 'meta' instructions about the unit of measure, as some sort of warning: *ratio*, *number*, *pure number*, *unitless number* are the most frequent, or to escape the problem the value is multiplied by 100 and presented as *percentage*. The problem with these practices is that they are not units of measure, but rather hints for the user of the data to investigate more thoroughly the indicator definition or look-up methodological metadata to have a sense of the context in which the number can be used. In a way, using special code-list items like \_X: 'Not allocated' would be more appropriate than the meta-instructions.

### 'Change of unit-of-measure' approach:

We still divide debt with reserves, but this time we recognise that the two quantities are commensurable, so the resulting quantity - exactly the same ratio as above - can be interpreted as the quantity in the numerator expressed in multiples of the quantity in the denominator.

$$\text{reserves} = 4000 \text{ USD} \Rightarrow \frac{1}{4000} \text{reserves} = 1 \text{ USD} \quad (4)$$

thus

$$\text{gross debt} = 5000 \text{ USD} = 5000 \cdot \frac{1}{4000} \text{reserves} = 125 \% \text{ of reserves} \quad (5)$$

In this approach the value of 1.25 (or 125%) is associated with the quantity in the numerator of the indicator in (3), and the quantity represented by the denominator becomes the new standard for value measurement (i.e. the new unit of measure). Notice that, when converting units of measure, it is the measured quantity that becomes the new unit of measure and not its original unit of measure (It is a recurrent mistake to carry the original unit of measure of the denominator, rather than the denominator itself into the new unit of measure). This highlights, how units of measures are arbitrarily chosen measured quantities, and within the same dimension (i.e., when commensurable) they can be used interchangeably once the proper conversion factors have been applied.

The second representation of the value 1.25 in most contexts is more informative and often analytically more useful than the presentation of a ratio with no unit of measure associated. The 'change of unit-of-measure' approach fixes the problem of missing unit of measure by maintaining the association of the measured quantity with the original

dimension, and hence it remains clear that the number – in our case – is associated with a *monetary value* and its size is proportional to the quantity in the denominator (the reserves). The quantity in the denominator is often chosen to normalise quantities and thus make them more directly comparable with other value quantities in the same economy. Caution must be taken though when comparisons are made across countries as the units of measure like 'percentage of reserves' are context dependent, and are meant to be different across countries<sup>[^2]</sup>, limiting additivity – however, often supported by economic theory they maintain a good level of comparability. It would be possible to remove context dependence of such units of measure, but the cost of it is to maintain a much larger set of units of measure: e.g., 'percentage of reserves of country X' for each country X.

## Pressure to use inconsistent variations

Because of the alternative presentation options for commensurable quantities and the subtle difference these options induce vis-a-vis the unambiguous rule for not commensurable quantities there are precedents and pressures to use inconsistent combinations of measure and units-of-measure. We have observed two types of deviations from a consistent unit of measure use.

The first variant is when data producers compute an indicator of non-commensurable quantities, yet the 'change of unit-of-measure' approach is forced, even though that would be only applicable to commensurable quantities.

For example, the measure 'GDP to population ratio' which should be associated with 'USD/person' is simply presented as with the measure 'GDP' and a clumsy unit of measure resembling 'USD/person'. This variation is clearly inconsistent and could lead to surprising results if taken literally and applied in computations that adhere to the rules of dimensional analysis. The appeal to use this representation is there because in a large database one would like to be able to align/pivot all variations (and derivations of GDP). Nonetheless it would be a better solution to introduce a dedicated dimension which captures the 'projection' basis which in this case would be 'population', in other cases could be 'households' or 'hours worked'. The measured quantity would emerge as a combination of two dimensions: GDP in the measure dimension and the projection base in the other dimension, and the correct unit of measure would apply to each combination. This would avoid tinkering with the unit of measure itself.

The careful reader might have already noticed that GDP is assumed to carry a wrong unit of measure (like in most statistical databases). This is because GDP being a measure of flow, i.e. value added generated over a period of time, should have units of measure that convey this e.g. USD/year or USD/quarter, rather than USD alone. The reason why – by convention – the unit of time is not marked in the unit of measure is that it is assumed to correspond to the frequency of the time-series in which the GDP figures are presented<sup>[^3]</sup>, or it is the well known frequency of the analysis. However this convention can lead to confusions. For example, annualised figures of GDP are sometimes used in a quarterly context. Typically, a debt-to-GDP ratio is calculated with annualised GDP even when working with quarterly data of debt and GDP, and users unaware of the convention would not know if the annual debt-to-GDP ratio is immediately comparable to the one calculated in a quarterly context (hint: it is). Such a mental gymnastics could be spared if the unit of measure for debt-to-GDP would be *years* or *quarters* as it follows from dimensional analysis.

The other compelling case to deviate from a consistent use of units of measure is opposite of the previous one. In this case the starting point is an indicator computed from commensurable quantities and the 'unitless number' representation for the measure, based often on the argument that this is the established convention, this is how users recognise the indicator. Take for example the 'unemployment rate' (literally it is the unemployed to labour force ratio) and in a consistent unit of measure allocation this is presented as *percentage*, however to give context and avoid misinterpretation we recommend associating it with *percentage of labour force*. This approach, while it seemingly kills two birds with one stone (familiar measure, informative unit of measure) can still cause problem when used in further computation and dimensional analysis is applied literally.

## Recommendations

1. Use consistent representations whenever possible
2. When calculating indicators of commensurable quantities, prefer the 'change of unit-of-measure' representation
3. When convention dictates the use of inconsistent representations provide ample, easily accessible metadata and clues for the user to be able to get back in the comfort of dimensional analysis for further derivations, or be able to remain in the 'quasi-consistent' world of the established convention.

<sup>[^1]</sup>: Indeed the guideline's perspective here is quite unusual and very utilitarian. In a classical approach we know in advance what we measure and the unit of measure follows. Here the numerical value emerges first and we architect the data model around it, and ponder the utility of two distinct combinations of 'measure - unit of measure pairs' that will go with the number, and set scopes of usage for it.

[^2]: The analogy may sound far fetched, but even in science we have limitations on how universal the units of measure can be. For example a metre in one inertial reference frame cannot be combined with a metre from another sufficiently incompatible reference frame.

[^3]: To bring a science analogy: this is as if the distance covered by a moving object would be plotted every second. Each individual measurement is a distance (in meters), but it also corresponds to the velocity of the object at that instant (in meters per second) as our measurement frequency exactly corresponds to the unit of measure of time.

## 2.3. Measuring the numerosity of sets

---

In the example above on dimensional analysis we have already touched upon the issue of measuring the size (numerosity) of a population (set of distinct elements), handling it in an intuitive way. This section extends the analysis to similar scenarios and generalises recommendations to measurement of the numerosity of various sets and the association of units of measures with such measurements.

As usual our starting point is to look for analogies in the the SI system for measurements. The SI system recognises the *mole* (symbol: mol) as the standard unit of measure for the amount of substance, which depending on the context could be molecules, atoms or sub-atomic particles.

The definition of the *mole* in SI system provides a good entry for measuring the numerosity of sets, or quantify the size of sets with varying number of discrete elements in them. However in the fields of economic and social statistics, the mole with its embedded unit-multiplier (the very large Avogadro number) is rarely a practical unit of measure, and the name of the associated dimension (i.e. amount of substance) is rarely a good descriptor of the type of statistical population (or more in general 'set of discrete elements') we would like to characterise. Hence the name of the section: measuring the numerosity of sets, and it is meant as a generalisation of measuring the amount of substance.

In an attempt to generalise the measurment of numerosities we revisit the very act of measurement and break it down so that parallels can be found between the common act of measurement over continuous quantities and the size measurement of sets with discrete, finite elements. Recall that, when measuring continuous quantities, we start by defining a standard, then we define ways of comparing the standard to the measured quantity (typically by 'recognising' the standard multiple times in the measured quantity), and finally counting how many times we were able to 'align' the standard with the measured quantity (without overlaps). So when moving from continuous quantities and their measurement to measuring discrete quantities we should construct a unit of measure that:

- can be well defined as a standard (i.e. is recognisable),
- the elements of the set can be 'aligned' or 'matched' to that standard
- and the distinct matches can be counted without overlaps

Following from this a typical statement of:

$$\text{Measure} = \{x\}[\text{unit of measure}] \quad (6)$$

for a numerosity measurement would take the shape of:

$$\text{A description of the set} = \{x\}[\text{an 'ideal' element of the set}] \quad (7)$$

For example:

$$\text{Population of Seborga} = 297 \text{ persons}$$

One outstanding question remains - how should the 'ideal' element of the set be described? When working with practical cases we often come across the problem of the possibility of assigning the characteristic element of the set in specific or generic terms. In the above case we could have opted to use 'inhabitants', 'persons' or something more generic e.g. 'entities'. They would all be valid units of measure according to the principles laid out above. How would we nonetheless decide which one to choose? Is it the approach of the SI system - to choose a very generic term - the one to follow? Perhaps not, or not in all circumstances. The best unit of measure should help the user of the data to understand the scope of the data, to clearly set out the limits of comparability of the data with similar measures and inform on the range of simple operations that can be immediately performed on the data.

Let's work further with the above example and analyse the implications of choosing more specific or less specific units of measure.

- **Precision**

If we were to choose 'inhabitants' as the unit of measure we would very clearly indicate to the reader which persons were counted in the population of Seborga, it would be obvious that someone born in Seborga but living abroad is not counted, as well as tourists passing-by would not be counted. To achieve the same level of precision with 'persons' as the unit of measure one would have to provide referential metadata to describe the measurement methods and the immediacy of that information would be lost. Going for 'entities' would introduce even more ambiguity, as the reader of the data would be unsure whether the population of Seborga measured in 'entities' would also include 'animals' or maybe 'enterprises' and other legal entities residing in Seborga. And even if the measure itself clarifies such questions, the user of the data would wonder whether additional measures intended to be regrouped or combined with the measure of Population of Seborga exist in the dataset or in the broader context.

- **Calculation scope**

Using specific units of measure like 'inhabitants' or 'tourists' allows the development of indicators (ratios) with units of measure of 'tourists/inhabitant', but would limit lumping them together, and deriving a quantity such as 'All people who set foot in Seborga'. If we would like to be able to construct such a quantity, 'persons' would be the appropriate choice of unit of measure, as it applies to both tourists and inhabitants, and when added together would follow the rules of dimensional analysis - i.e. the requirement to only add or subtract quantities with shared units of measure.

## Recommendations

1. When measuring numerosity of sets, use a description of the set as the measured quantity and use a description of the ideal element of the set as the unit of measure.
2. Of the many options for the 'ideal element' of the set, choose the one that best fits the intended precision and calculation scope.
3. When in doubt, contrary to conventions in science and engineering, lean towards more specific units of measure. It gives more precision, and it is easier for the end-user to generalise the scope when needed. Do not worry about the potential redundancy that might emerge between the *measure* and *unit of measure* wording.

## 3. Data-warehousing - representation of units of measure through multiple concepts

In a data-warehousing context data is structured in multidimensional cubes<sup>[4]</sup>. SDMX (statistical data and metadata exchange) standard provides a rich language to describe such cubes. <sup>[5]</sup> In this section of the guidelines we discuss the implications of multidimensionality compared to the simpler representation set out in the referenced standards for Units of Measure.

We would like to bring together the representation in  $Measure = \{x\}[\text{Unit of measure}]$  in (6) with the typical arrangement of data in a cube (where data are identified as points at certain co-ordinates in the multidimensional cube and where in addition to the measured value  $\{x\}$ , attributes are stored):

$$(Dimension_1, \dots, Dimension_n) = x(Attribute_1, \dots, Attribute_k) \quad (8)$$

The aim is that both the *Measure* and the *Unit of measure* emerges as a subset of Dimensions and Attributes, in an unambiguous and complete way. That is, each component (Dimension or Attribute) is either part of the *Measure* or part of the *Unit of measure* and the association of the component is consistent throughout a cube. This exercise, though it seems trivial, in practice can cause surprising challenges.

### 3.1. Should units of measure be modelled as dimensions or as attributes?

To best answer this question one should ask why would one choose a component to be a dimension or an attribute in a Data Structure Definition in general.

Based on the SDMX guidelines the concepts which both identify and describe data should be treated as dimensions. They are necessary to understand the meaning of data. To recognize whether a concept can be a dimension or not, it is important to understand the role it takes in describing the data.

One should consider which of the SDMX concepts are used to identify unambiguously an observation. Each of these concepts then should be considered as dimensions. An attribute on the other hand is a concept that only describes the data, but it is not used to identify the data. It can be mandatory or conditional, mandatory attributes are necessary to correctly understand the data and they should always be included with the data, whereas conditional attributes most often only represent additional information about the data – or, in some cases, are applicable only for a slice of the

cube in which they are defined (e.g. a base period applies to indexes, but is not relevant for current price data expressed in currency terms).

The rule described in the previous paragraph should be the main one, but occasionally, driven by intended uses attributes can be 'upgraded' to dimensions. This is typically the case when tools with which we manipulate the cubes have richer behaviours associated to dimensions than to attributes (e.g. filtering, pivoting only available to dimensions). If we perceive that attributes could improve data manipulations they can be upgraded to dimensions, although these upgrades should happen sparingly, as they come with a cost – they make the data identifiers longer and hence more costly to use (e.g. if a concept organised in a hierarchy is accompanied by an attribute that captures the level information – the level information as a dimension can be used for easy data slicing, but it also means that the level information has to be provided every time we reference granular data, even though the underlying hierarchical code already contains everything we need to identify them).

How does this translate for *Units of measure*? As it was pre-empted in the mission statement below (8), the general modelling principles do not offer a prior guidance to whether the data-structure should treat the units of measure as dimensions or attributes. There are scenarios when units of measure are just attributes (typically datasets with few time-series, sitting sparsely in a large conceptual space) but often units of measures are needed as dimensions, or indeed, when units of measures are composed of multiple dimensions and attributes (e.g. economic data where the same quantity is expressed in multiple ways: different currencies, different price base conventions, or normalised in various ways). Of course, in the latter cases it is always possible to construct an additional attribute that captures the unit of measure as a whole, but this would come at cost of redundancy in the data model, moreover – it would not really change the association of certain dimensions from the *Unit of measure* to the *Measure* in the metamodel of (6).

[^4]: Data is offered in multidimensional cubes rather than in the simple form because this more structured form facilitates analysis in at least three ways: a richer context of the data is revealed by the component structure; the atomic organisation facilitates filtering, slicing and pivoting the data; and connections across cubes can be more easily established (as long as data are modelled in harmonised ways, e.g. data referring to the same country or the same product can be connected across cubes).

[^5]: Throughout this document we tend to use a language that sometimes shifts between an abstract mathematical language and the concrete language of the SDMX standard. The reader should observe carefully the references to objects (artefacts) described in the information model of the standard (e.g. cubes are roughly equivalent to DSDs, 'Concepts' are abstract manifestations of 'Components' of a DSD, which themselves can play a role of a 'Dimension' or an 'Attribute').

## 3.2. Concepts to consider when assigning units of measure

---

In some statistical domains (e.g., Economy) the units of measure tend to be determined by multiple concepts. For GDP, for example the narrowly defined Unit of measure could be "National currency" or "US Dollar", but other concepts may also be necessary to uniquely pin down the full unit of measure (and provide an accurate unit of measure label). Price base details may be needed to distinguish between current price or constant price measurement variants of GDP, adjustment may be needed to differentiate seasonally adjusted and non-seasonally adjusted data, base years are needed for indexes, conversion-styles are needed when currencies are translated into each other and so on.

If we put all the concepts that are relevant to measurement into a single unit of measure label, the narrowly defined US dollars, may grow into something like:

```
US dollars, constant prices, constant PPPs, reference year 2015, millions
or
US dollars, current prices, current PPPs, seasonally adjusted
```

Moreover, having units of measure with all the possible combinations of the included components in a single dimension, and accordingly represented by a single code-list would be hard to manage. Therefore, the general recommendation is to split the complex units of measure into separate dimensions and attributes in the model whenever possible. Some of the SDMX tools are already capable of assembling composite units of measures from multiple dimensions and attributes. Having said that, if there is a need for editorial control that goes beyond simple concatenation rules – and attribute can always be added to the data-model.

Here is how the above example US dollars, constant prices, constant PPPs, reference year 2015, millions could be decomposed into several unit of measure components:

Narrow unit of measure: US dollar  
Price base: Constant prices  
Conversion type: PPP converted  
Base period: 2015  
Unit multiplier: Millions

Table 1 presents a list of concepts associated with units of measure and their recommended representations. These concepts often have a considerable impact on measurement. Sometimes they literally become part of the unit of measure, other times they just shape the narrowly defined unit of measure.

**Table 1** Unit measure related concepts and their representation

Concept ID	Concept name	Concept representation	Concept description	Examples
UNIT_MEASURE	Unit of measure	CL_UNIT_MEASURE	Unit in which the data values are expressed. Often this is what we refer to as the narrow unit of measure.	Persons, National currency, Hours per day
CURRENCY	Currency	CL_CURRENCY	Currency, could represent the denomination or the valuation of the measure	Euro, US Dollar
UNIT_MULT	Unit multiplier	SDMX:CL_UNIT_MULT(1.1)	A multiplier to calculate the actual value in the narrow unit of measure.	Units, Thousands, Millions
TRANSFORMATION	Transformation	SDMX:CL_TIMETRANS(1.0)	Time-related operation performed on a time series, involving only observations of that time series.	Index, Growth rate
ADJUSTMENT	Adjustment	SDMX:CL_SEASONAL_ADJUST(1.0)	Procedures to decompose series into trend, cycle, seasonal and outlier components	Seasonally adjusted, Trend
CONVERSION_TYPE	Conversion type	CL_CONVERSION_TYPE	Performed currency conversion	PPP converted, Exchange rate converted
PRICE_BASE	Price base	CL_PRICES	Indicates the set of prices used for the valuation of the measure, often accompanied by base period information.	Current, Constant, Chain linked
BASE_PER	Base period	ObservationalTimePeriodType	Period of time used as the base of an index number or to which a constant series refers.	2010, 2015
BASE_REF_AREA	Base reference area	CL_AREA	Reference area used as the base of an index number.	EU, OECD, US

### (Narrow) Unit of measure

The UNIT\_MEASURE concept in Table 1 is often the only component to describe the full Unit of measure, but when other dimensions or attributes contribute to the full unit of measure, this concept is meant to capture the most important, core part of it and we often refer to it as the 'narrow' unit of measure in the data model. The simplest but most frequently used extension is the addition of a Unit multiplier in which case the 'full' unit of measure: USD, millions is represented as a combination of two concepts the 'narrow' unit of measure: USD and the unit multiplier: millions.



## Currency

Currency codes are often part of the narrow unit of measure code-list, yet there are two well identified modelling use cases when the separate representation of currency (concept and codelist) is needed.

1. When currency is used as a denomination currency (and not as a valuation currency) it has no impact on the unit of measure
2. If the currency is used for valuation, there are two distinct data modelling variations.

One modelling variant relies on a dedicated currency concept and code-list. It offers a richer structure and it is particularly useful when the cube contains quantities in both national currency and converted into a single currency to facilitate cross country comparisons. When the data is in national currency, it should be represented such that:

```
UNIT_MEASURE (dimension/attribute) = 'National currency' and  
CURRENCY (attribute) = <a specific currency> (e.g. AUD)
```

whereas when the data is a quantity converted to USD its representation is as follows:

```
UNIT_MEASURE (dimension/attribute) = 'USD converted' and  
CURRENCY (attribute) = USD
```

The other modelling variant relies only on the Unit of measure concept to encode currencies, e.g. applicable when the data is only in national currency. The narrow unit of measure is set to a specific currency (note that all currencies are also listed in the CL\_UNIT\_MEASURE code-list).

The CL\_CURRENCY code list contains the list of existing and former currencies listed in the ISO 4217 standard. For the 'currency denomination' use case it may contain references to currency pools (e.g. Major currencies, EU currencies) which are strictly forbidden for units of measures. Therefore, in the design of the Unit measure code-list such currency pools and non-specific currencies should be avoided.

## Unit multiplier

Unit multipliers are used for readability, and they also double as an implicit measurement precision indicator (i.e., an economic value presented in millions is unlikely to be precise to the cent). In the SDMX code-list for unit multipliers the codes are powers of 10 needed to construct the actual multiplier. When the unit multiplier is 1, it can be omitted.

Some unit multipliers are different. The 'Greeks' are often added as variations of the unit of measure into the unit of measure code-list: e.g. grams, kilograms, hectograms, but their English equivalents (thousands, hundreds) are in the dedicated unit multiplier code-list. Nonetheless, the different variants can be freely combined, e. g. kilograms, millions.

Negative unit multipliers (percent, per thousand, per million), frequent with derived indicators in social statistics, also tend to be directly listed in the unit of measure code-list rather than composed from the dedicated unit multiplier code-list. This choice is mostly motivated by the readability issues observed when concatenating the narrow unit of measure with negative unit multipliers (e.g., concatenated: 'per birth, per thousand' vs. direct: 'per thousand births' ).

## Transformation

Using a transformation dimension in a data model is challenging, especially for the proper unit of measure allocation. The problem is that some transformations alter the measure as well as the unit of measure, and hence cannot be simply associated fully with one or the other. An index calculation (simple rebasing) does not change the measured quantity (e.g. a monetary value stays a monetary value, just measured differently), whereas growth rate calculation produces a whole new measure – to take a science analogy an index calculation is similar to changing the way a distance is measured, whereas a growth rate over time is similar to a calculation of a distance over time, i.e. a measurement of speed.

The transformation dimension helps to connect nicely a key measured quantity and its transformations, but the resulting data-models are typically more difficult to interpret. If the measures are altered to capture the differences in nature between the original data and their transformations (the easy slicing or pivotability of the data model disappears).

Additionally, there is the dilemma how to assign units of measures to measures: pre-transformation or post-transformation. The pre-transformation approach is informationally more complete (but then the narrow unit of measure is not the proper unit of measure to associate with the measured number). The post-transformation approach is aligned with the presented data, but it may be that the model becomes hard to read and needs additional attributes or dimensions to disambiguate or make the data fully interpretable.

In general, especially when designing cubes for dissemination to wider audiences avoid the transformation dimension. If a transformation dimension is included nonetheless, represent the post-transformation unit of measure in the UNIT\_MEASURE concept (i.e. index for indexes or indeed 'percent of base period value', and 'percent per annum' for a growth rate calculated over a year).

## Adjustment

This is typically used for a special class of transformations that emerge as part of seasonal adjustment process. Time-series, when looked at as combinations of cycles of various cycle length, are composed of an underlying trend (fluctuations of 10 year and above), a cycle (fluctuations between 2 to 10 years, definitions vary), seasonal variations (less than one year) and outliers of various shapes (level shifts, transitory changes, spikes). The seasonal adjustment can be interpreted as a different measured quantity, but, as it is often the case, it could be also interpreted as the same measured quantity but measured with a measurement standard that changes in time to compensate for seasonal variations that obscure underlying twists and turns. As a result of the latter interpretation, seasonal adjustment is often presented as part of the unit of measure (e.g. USD, seasonally adjusted).

## Price base and Conversion type

Both *Price base* and *Conversion type* are similar in nature to seasonal adjustment, but their impact on the 'metre rod' is more straightforward to justify. When measuring value across time, we realise that the unit of value in one period of time has potentially a different value in another period of time – it does not reflect the same value anymore. Typically, if all prices of all things valuable doubled overnight, then the pre-doubling currency unit would represent twice the value than the post-doubling currency unit, inflation eroded the value of the currency unit as compared to other valuables.

Therefore, in macroeconomic studies comparisons are often made with a unit of measure that is corrected for the impact of time. Such time correction can be made for example by asserting the value that would be observed if prices did not change (only quantities). For example 'constant price USD' measures the changes in value, after correcting for across the board price changes.

The conversion type is used when comparisons are made spatially, and specifically between areas of different currency units. Converting between two currency units can be done based on purchasing power parities or based on observable exchange rates. As these can have an impact on the measurement outcome the same way as time corrections, their specification next to the underlying reference currency is crucial.

Sometimes the conversion type is specified in a dedicated dimension, but often, the data space is compact enough to introduce the conversion type variants, for the few reference currencies (USD and EUR) in the 'narrow' unit of measure code-list.

## Base period and base reference area

These concepts are important when data has been expressed as a factor of a quantity observable in a specific time period (base period) or in an equivalent quantity in a different geographic or economic reference area (base reference area). There are two equivalent ways to assign a unit of measure when base periods or base reference areas have a role in the data model. When the measure is expressed as a multiple of its value at a given time period, we could either have: Index, 2015 as the 'full' unit of measure, but equivalently we could also say that the unit of measure is 'Percentage of the measured quantity in 2015'. Depending on the audience both approaches can be used, although a consistence across at least a data-warehouse is desirable. For time-reference quantities the former pattern is more common, as it allows for an easy combination of the unit measure and base period (where base periods can be numerous). In the case of base reference areas, the latter pattern is more popular e.g. 'Percentage of the measured quantity in OECD', and can be expressed in a compact form only requiring the unit of measure concept.

## 3.3. Coding conventions and a typology of units of measures with examples

---

In this section we present conventions and recommendations for the narrow unit of measure, as described in the previous section. According to the SDMX Glossary [5], the cross-domain concept for the unit of measure is `UNIT_MEASURE` "Unit of measure" which defines the unit in which the data values are expressed. The recommended representation for this concept is the code list `CL_UNIT_MEASURE`.

In some global Data Structure Definitions (DSDs), the concept `UNIT` is used instead of `UNIT_MEASURE` with the code list `CL_UNIT` instead of `CL_UNIT_MEASURE`. When creating structures, it is recommended to align with the SDMX Glossary recommendations.

### 3.3.1. Adding additional information through annotations

To communicate symbols for units of measures, it is important to use the correct symbols, follow the standard conventions, and avoid ambiguity. Since the SDMX technical standard is case insensitive by nature, we are often misaligned when representing the official symbols/codes used by existing standards. For example, the symbol for metre is *m*, not *M*, which stands for mega, a prefix that means million. The symbol for second is *s*, not *sec*, which is an abbreviation, not a symbol. The symbol for kilogram is *kg*, not *Kg* or *kG*, which are incorrect capitalizations. Some computer systems or programming languages still have the requirement of case insensitivity and some humans who are not familiar with SI units tend to confuse upper and lower case or can not interpret the difference in upper and lower case correctly.

For this reason, the case insensitive symbols are defined. Although the Unified Code for Units of Measure does not encourage use of case insensitive symbols where not absolutely necessary, in some circumstances the case insensitive representation may be the greatest common denominator. Thus, some systems that can handle case sensitivity may end up using case insensitive symbols in order to communicate with less powerful systems.

To avoid ambiguity, it is also important to use "Code item" annotations to add alternative representation of symbols. As an example, we should add an annotation to reference the official UCUM. When referring to kilograms, we could have the code id set to "KG" within our Units of Measure codelist. Since the official UCUM representation is "kg", we should add the following annotation to the "Code item".

```
Where to attach: Code item
Use Case: The official representation of the Unified Code for Units of Measure (UCUM).
Annotation Type: UCUM_CS_CODE
Annotation Title: kg
```

### 3.3.2. Types of units of measure, coding patterns, labelling of UoMs

Units of measure can be classified into distinct categories depending on their origin, derivation and intended usage. Here are the most common categories and suggestions of how these could be used and coded.

**Scientific units of measure** These are the units of measure coming from standards like SI, ISO/IEC 80000, UCUM and others. It is the most stable part of the unit of measure code list as there are agreed codes and names that change very rarely.

*Suggested coding pattern:* use agreed scientific codes (e.g. H for Hours, KG for kilograms). In order not to introduce additional concepts in the models, these codes can use prefixes (milli, micro, kilo, mega etc).

*Code examples:*

```
H: Hours
HA: Hectares
HL: Hectolitres
KCAL: Kilocalories
KG: Kilogramme
KJ: Kilojoules
```

**Itemisable units of measure** These are the units of measure that can be used to count the items in discrete quantities. These can be more generic, e.g., Persons, or be more specific, e.g., Children, Students, Investors. However, it is not recommended to use units of measure like Number. A good unit of measure for numerosity of sets should answer to a question "number of what?".

*Suggested coding pattern:* use 2-4 character shortened codes, remove vowels, make sure that there is no duplication with ISO currency codes.

*Code examples:*

PAT: Patients  
PR: Permits  
PS: Persons  
RCP: Recipients  
RG: Registrations  
RO: Rooms  
RW: Registered workplaces  
SB: Subscriptions  
SC: Schools  
ST: Students

**Currencies** The list of currencies is based on the ISO 4217 standard. It includes currently existing and former currencies, funds, precious metals etc, also codes like "National currency" or "Reported currency" may be needed as unit of measure.

*Suggested coding pattern:* in the case of national currencies, the first two letters of the alpha code are the two letters of the ISO 3166-1 alpha-2 country code and the third is usually the initial of the currency's main unit. XDC for national currency.

*Code examples:*

TRY: Turkish lira  
TTD: Trinidad and Tobago dollar  
TWD: New Taiwan dollar  
TZS: Tanzanian shilling  
UAH: Hryvnia  
UGX: Uganda shilling  
USD: US dollar

**Derived units of measure** These are the units of measure that are usually formed as a result of combining currency and scientific/itemisable units of measure (by applying dimensional analysis), e.g., National currency per person, Euros per hour, US dollars per tonne, Days per week, Kilogrammes per person.

*Suggested coding pattern:* use the same scientific/itemisable/currency codes and "\_" as a separator.

*Code examples:*

H\_D: Hours per day  
H\_PS: Hours per person  
H\_WK: Hours per week  
H\_WK\_PS: Hours per week per person  
H\_WR: Hours per worker

## Percentages

When defining percentages, it is recommended to be specific (particularly in the context of 'change of units-of-measure' style modelling), that is, instead of "Percentage", use "Percentage of (something)", e.g. Percentage of GDP, Percentage of population, Percentage of labour force etc. Percentages can be even more specific, adding precision of the selected items in them, e.g., Percentage of women aged 15-49 years, Percentage of intermediate consumption in the same product, Percentage of working age population in the same subgroup.

*Suggested coding pattern:* PT\_(code of indicator/measure)\_(codes of relevant breakdowns)

*Code examples:*

PT\_POP: Percentage of population  
 PT\_POP\_AGE: Percentage of population in the same age  
 PT\_POP\_SEX\_AGE: Percentage of population in the same sex and age  
 PT\_POP\_SUB: Percentage of population in the same subgroup  
 PT\_POP\_Y\_LT1: Percentage of children aged less than 1 years  
 PT\_POP\_Y\_GE15: Percentage of population aged 15 years or over  
 PT\_POP\_Y15T64: Percentage of population aged 15-64 years

Besides the 'Percentage of *something*' pattern it is worth mentioning a few related codes and units of measure. To complement growth rates in time series, or interest rates and yields it is recommended to use 'Percent per annum' (PA) or 'Percent per period' (PP) as unit of measure. Their close relative 'Percentage change' is not recommended as it withholds important information about the applicable timeframe – and it rarely adds to the information in the Measure (or the Transformation dimension). Additionally, 'Percentage points' (PD) are often used in contexts where additive manipulations are carried out on quantities already expressed as percentages (e.g. a difference of two interest rates) or a decomposition of a percentage term (e.g. various expenditure component's contribution to GDP growth).

**Factor of ...** These units of measure are meant to replace the unitless units of measure like Ratio or Pure number (that are not valid units of measure). This kind of units of measure can be used for ratios when one figure is divided by another, and the result basically shows how many times one figure is greater than another, for example, Factor of GDP. The 'Factor of ...' pattern is similar to the 'Percentage of ...' pattern. The only distinction is that for the latter ratios are multiplied by 100.

Government debt = 1.19 factor of GDP, or,  
 Government debt = 119 percent of GDP

*Suggested coding pattern:* FCTR\_(code of indicator/measure)\_(codes of relevant breakdowns)

*Code examples:*

FCTR\_D1: Factor of decile 1  
 FCTR\_B1GQ: Factor of GDP

**Other, special units of measure** Special codes like \_X: "Unspecified" or \_X: "Not allocated" may be used in some cases as units of measure (particularly in the context of describing ratios quantities with the same unit of measure), but \_Z: "Not applicable" should be avoided.

A different class of special units of measure is tied to measurements on nominal or ordinal scales. The measurements described above have almost exclusively been done on continuous, interval scales (meaning that addition/subtraction and multiplication/division between measured quantities are meaningful), perhaps the exception of the class of units of measure referred to as *itemisable*. These latter cannot be regarded as continuous scales and hence division does not always yield easily interpretable outcomes, but for most practical use cases they behave well in dimensional analysis. The ordinal and nominal scale measurements however represent a coarser *measurement* of reality, and they at best only convey order between measurements, but addition and multiplication cannot be used and thus the applicability of dimensional analysis is limited. Examples of such scales are: Pantone scale for color, Likert scales used in questionnaires (1 to 5, or other ordered set of choices measuring intensity), Beaufort scale for wind intensity.

Ultimately some measurements, even though are not strictly speaking related to ordinal scale measurements, may be treated in a special way just because the transformations that were applied to the measurements result in non-linear scales, or capped/bounded scales which limit the further additive multiplicative extensions of the measured quantities. Presenting scale information for such measurements may be more relevant than describing a *standard*. Examples could be: the Richter scale for earthquakes, pH-scale for acidity.

For these measurements the recommendation is to provide *units of measure* that describe the scale on which the measurement has occurred. These scale descriptors will be incorrect units of measure in the strict sense, nonetheless they will serve the same purpose in guiding the data-user: help understand the limits of comparability with other quantities and the scopes of computation in which these numbers can meaningfully take part.

*Suggested coding pattern:* SCL\_(scale bounds or common abbreviation)

*Code examples:*

SCL\_1T5: 1-5 scale  
SCL\_0T100: 0-100 scale  
SCL\_PH: pH scale

Please note that the various codelist categories can have varying degree of volatility, some are fairly static others change often or are strongly context dependent. The static code lists are often associated with reference data from mature frameworks such as scientific unit of measures and currencies. For these static code lists, the SDMX – Statistical Working Group (SWG) will manage and publish the appropriate cross-domain code lists. The more dynamic code lists include categories that frequently change over time, and from one domain to the other requiring additions to reflect particularities in statistical data products. Examples of dynamic code lists include derived units of measure, percentages, and factors of. For these dynamic code lists, the management responsibilities will lie with the various business domains, however it is recommended to adopt the coding patterns described in the guidelines to maintain broad convergence.