Statistical Working Group
Statistical Guidelines

_____

Guidelines for modelling Units of measure in SDMX
Version 1.0 - 30/11/2025

_____

DOCUMENT HISTORY

Version	Date	Comment
1.0	2025	Initial version

1. Introduction

The role of units of measure in statistical data modelling is critical for ensuring consistency, accuracy, and interpretability. Measurement units provide a standardized way to quantify data, enabling immediate assessment of comparability and defining the scope for computations.

Accurate and consistent unit conversions are essential for maintaining data integrity and enabling meaningful analysis when comparing statistical indicators. This underscores the importance of guidelines for Unit of Measure (UoM) metadata modelling. These guidelines establish a framework for organising and standardising how units of measure are represented and managed within SDMX data models.

The Statistical Data and Metadata Exchange (SDMX) consortium has developed tailored guidelines to promote harmonized cross-domain concepts and terminology, improving the efficiency of data exchange. These guidelines for modelling units of measure, along with the planned work on Unit of measure associated code-lists fit into this series of harmonisation efforts.

These content-oriented guidelines are designed for statistical data modelers working with multidimensional data, primarily in SDMX, but they also extend to data-warehousing applications in general. The patterns proposed are rooted in well-established international standards in science and engineering.

Beyond the reference to well-established unit of measure standards, there are three novelties addressed in this guideline: 1. it describes dimensions and their corresponding *units of measures* that are typical to economics and social statistics, 2. it describes two alternative but equivalent representation variants of ratios of commensurable quantities 3. 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 goal is to preserve the essence of the existing standards while integrating the new dimensions, that is, facilitate comparability of the data across a wide spectrum of data sources and clarify the computational contexts in which the data can be used.

Ratios of commensurable quantities are often described as unitless numbers. Nonetheless this approach in a data-warehousing context can lead to an inflation of hard to compare and interpret numbers. An alternative but equivalent representation, dubbed as the 'change of unit of measure' approach is presented, which better informs data users about comparability of the data.

The third 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 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 a multidimensional cube. 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).

By adhering to these guidelines, organisations can achieve greater data consistency, improve interoperability, and enhance the overall quality of their data.

2. Alignment to existing international standards for Units of Measure

2.1 Reference frameworks

The guidelines put forward in this document do not operate in a vacuum. Instead, they build directly on well-established standards, classifications, and ontologies for units of measure. Any deviations from these existing frameworks are few and will be thoroughly explained in later sections.

At the core of these frameworks is the **International System of Units (SI)** [SI], maintained by the Bureau International des Poids et Mesures (BIPM). Widely adopted and promoted by metrology institutes worldwide, the SI system defines the fundamental concepts of quantity and unit. It establishes seven base quantities and their units, which serve as the building blocks for all other units through simple multiplication and division.

Beyond the seven core dimensions, the SI system also accounts for quantities that represent "number or entities". The system assigns a unit of 1 - a convention that will be discussed in section 2.5. The SI system is generally a closed system , with the ambition to create global convergence and standardization, achieved by normatively defining the base units for each dimension. This helps reduce errors associated with unit conversions, which are less needed with full standardisation as opposed to necessary conversions when competing units are in use.

Despite its benefits, the SI system's rigid, prescriptive nature has led to the development of other, more flexible unit of measure standards. This document considers two in particular: **ISO:80000-1** [ISO] and the **Unified Code for Units of Measure (UCUM)** [UC]. Both standards are largely aligned with the SI but differ by acknowledging the arbitrary nature of base unit choices and accepting a wider range of commonly used units.

While these standards offer an abstract view of dimensional systems, they do not provide sufficient guidance for measurements in fields like economics and social sciences. To fill this gap, this document introduces a base quantity type for "monetary value" and proposes a different approach for measuring the "numerosity" of finite sets ("number of entities" issue), diverging from the methods of the other standards.

2.2 Transposing Unit of Measure standards to a digital world

The digital era has seen numerous disconnected efforts to create unit of measure recommendations. In their simplest forms, these have been domain-specific classifications and code-lists, such as those from UN/CEFACT [UNC] for the trade and electronic business community. More complete digital artifacts, like ontologies, have also emerged, some related to the frameworks in Section 2.1, while others have grown within communities like NASA and UN/CEFACT. Keil and Schindler provide a comprehensive comparison of these ontologies in their *Semantic Web Journal* paper [KS]. The SI system itself has also established its own digital framework at https://si-digital-framework.org/SI?lang=en.

While this document does not provide a definitive list of units of measure, it outlines best practices and patterns for consistent identification and application, particularly in domains outside the scope of the referenced frameworks. The SDMX Statistical Working Group's Unit of Measure Task Group plans to design code-lists for various classes of units, but it has no intention of developing new ontologies. Instead, its design efforts will focus on granularly annotating and referencing corresponding concepts in established ontologies and in the SI reference point, ensuring alignment and interoperability.

The **DRUM** (Digital Representation of Units of Measurement) task group, under the auspices of the Committee on Data (CODATA) of the International Science Council, has the core mandate to

promote the FAIR principles (Findable, Accessible, Interoperable, and Reusable) for scientific data by ensuring units of measurement are not "second-class citizens". Its primary goals are to work with international scientific unions to raise awareness and provide guidance on implementing digital unit representations. This includes developing recommendations for machine-actionable encoding of units, creating guidelines for annotating data, and defining best practices.

2.3 Using dimensional analysis

In engineering and science - as a part of the mentioned standards [1] - dimensional analysis is systematically used to determine units of measures of measured, estimated and derived quantities. Dimensional analysis comprises of the analysis of quantities to identify their 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 quantity is expressed as multiples of the unit of measure. Additionally, dimensional analysis sets out the rules to working with 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 are possible between quantities, with the resulting quantity carrying the unit of measure that results from the same operations carried out on the units of measure, using the conventional rules of exponentiation. E.g. when computing a speed change over a period, the units of the resulting derived quantity will be: $\frac{m/s}{s} = m/s^2$

- 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 quantities (measurements, estimates and derivations), 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, 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 and avoids conversion and interpretation errors.

---

[1] e.g. UCUM section 3 - Semantics; SI Brochure 9, section 2.3.3 - Dimensions and quantities.

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.

2.4 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 quantities, by computing ratios of non-commensurable quantities :

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

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

These are all straightforward cases, following from the 3rd bullet-point on dimensional analysis, mentioned above in 2.3. The measured quantities are 'Debt to population ratio' and 'Reserves to population ratio' respectively, and in both cases the associated unit of measure is $\text{USD}/person$.

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 figure[2].

2.4.1 'Unitless number' approach:
$$\text{Debt to reserves ratio} = \frac{\text{gross debt}}{\text{reserves}} = \frac{5000 \text{ USD}}{4000 \text{ USD}} = 1.25 \text{ USD}^0 \qquad (3)$$

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 arithmetic $\text{USD}/\text{USD} = \text{USD}^0 \overset{\text{def}}{\cong} 1$.

---

[2] 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.

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 SDMX, using special code-list items like _X: 'Not allocated' would be more appropriate than the meta-instructions.

2.4.2 '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 - the same ratio as above - can be interpreted as the quantity in the numerator expressed in multiples of the quantity in the denominator.

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

thus

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

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.

2.4.3 Contextuality of units of measure

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 the 'missing' unit of measure by maintaining an association between the measured quantity and its original dimension (or quantity kind). In our example the number is expressing the size of a *monetary value*, and it represents a multiple of the reserves.

In certain systems notable quantities emerge as easy to interpret reference quantities. For example, in the solar system the average Earth-Sun distance (an astronomical unit) is such a quantity. Similarly, within an economy the size of reserves or the GDP can play such a role. They can be used to express the magnitude of other quantities representing monetary value 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 as reserves tend to be different across countries, additivity is not guaranteed for quantities of different countries – however, often supported by economic theory quantities expressed as '% of reserves' maintain some level of comparability (but still, not as direct comparisons of absolute magnitudes). 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.

In an ideal world units should not be context dependent, however in practice they often are. The units of the SI-system (its most recent definitions especially) relate to universal constants of nature, and hence they make the units of measure based on those definitions at first glance universal. Nonetheless at closer inspection, even base units like the metre have an implicit context. For example, a metre in one inertial reference frame cannot be additively combined or directly compared with a metre from another sufficiently incompatible reference frame.

From a contextuality point of view the unitless number approach yields the least precise, most contextual unit of measure. Assessing the magnitude is entirely left to the context. One must analyse the quantity kinds involved, the calculation methodology to be able to assess limits of comparability and additivity. A change of unit of measure approach offers the possibility to be more precise, less ambiguous, and delimit better the usage scope of the numbers. Nonetheless several options present themselves in the formulation of the units of measure, and they come with more or less unit information left for the broader context. E.g. in the example above, we have seen that the unit of measure can be '% of reserves' but could also be '% of reserves of country x' or '% of reserves of country x in year y'. The latter is the least contextual, it best guarantees that no accidental comparisons and additions are made with numbers expressed in different units.

2.4.4 The analysis of three common cases for ratios of commensurable quantities
We have seen that a good unit of measure leaves little to the context. Nonetheless, when formulating recommendations, this document balances multiple desirable properties of the unit of measure system:

- Activate the dimensional analysis benefits (set bounds of comparability and additivity), this corresponds to low contextuality, or precise and unambiguous formulation

- Be helpful in avoiding metadata lookups

- Do this without exploding the number of units of measure

There are three scenarios for ratios of commensurable quantities that do not mix very well together (especially when the 'unitless number' approach blends them together).

1. Dividing many quantities in a system with the same reference quantity

2. Calculating growth rates (especially for time series)

3. Calculating shares and proportions in sub-populations

Whenever a quantity is systematically used as a denominator to compute ratios with many other commensurable quantities (e.g. dividing all economic transactions in a country with GDP) there is a

strong appeal to regard these operations as a change of unit of measure. As with the *reserves* example above, the context is usually a country, and within the country additivity is maintained.

The second case of '%' calculations are less obviously tilting towards the change of unit of measure approach. *Growth rates* with the 'change of unit of measure approach' would yield '% of the base period value' and given that the base period value changes observation by observation, this unit of measure would be extremely contextual, and the context would be practically limited to a single observation. Thus, it would not offer any help setting bounds to additivity. The more practical 'unitless number' approach could be followed here, nonetheless, as it is typical in science and engineering with measurements that represent changes over time the recommendation is to use 'percentage per annum' or 'percentage per quarter' for growth rates. Growth rates computed over the same period are somewhat comparable e.g. can be presented on the same axis of a chart, but caution should be exercised with addition.

The third case *proportions* in various breakdown groups - e.g. proportion of unemployed to the total population in various age ranges - is the least obvious. On one hand the 'change of unit of measure' approach would lead to numerous units of measures of the form: '% of youth population', '% of 35 to 44 year population', etc., one unit of measure for each age range in our example. The 'unitless number' approach as usual would only assign '%' to all proportions, not just to the one in the example, but really to all proportions and shares of anything ever calculated. A contextual and imprecise form of unit of measure, midway between the two approaches could allocate 'percentage of same <subgroup>' for proportions. This approach can be useful within a data warehouse, especially when data is offered in a multitude of breakdowns, but proportions are only calculated with respect to a specific breakdown. In the example above population data may be available by age and gender, yet proportions of unemployed may be computed only by age, in which case, a helpful unit of measure would be 'percentage of same age population'.

We believe that the recommendations to articulate the units of measures for percentages a bit more than simply '%' helps reduce the need for metadata lookups, makes it possible to distinguish at least the three distinct cases mentioned above (reference quantities, growth rates, shares/proportions), allows comparability according to dimensional analysis and in the first case it even works for additivity rules (within the implicit context). The patterns here, while not the most economical, keep the number of variations in check - although not so much that a prescriptive list of all such units of measures could be provided in advance. A more compressed (%-only) system however would excel at the third requirement but would require major compromises for the first two requirements.

2.4.5 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 quantity 'GDP to population ratio' which should be associated with 'USD per person' in the global DSD for national accounts is simply presented with the quantity 'GDP' and a clumsy unit of measure resembling 'USD/population' (literally 'USD; ratio to total population'), shifting the description of the quantity into the unit of measure. 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 (USD per person, USD per household, USD per hour respectively). This would avoid tinkering with the unit of measure itself.

The other compelling case to deviate from a consistent use of units of measure is the 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. The choice of 'unitless number' representation is 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 it can be associated 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.

2.4.6 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.

2.5 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.

Our starting point is to look for analogies in the SI system for measurements. As already mentioned in point 2.1 the SI system deals with numerosity of sets in two distinct ways ('number of entities' and 'amount of substance'). The 'amount of substance' is recognised as one of the base dimensions of

the SI system, with the *mole* (symbol: mol) as the standard unit of measure, which depending on the context could be molecules, atoms or sub-atomic particles. Interestingly, for any other countable set no dimension is identified, but the convention of associating unit '1' with any other 'number of entities' is declared.

The definition of the *mole* in SI system provides a good analogy for measuring the numerosity of sets or quantifying 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, or an alternative way of dealing with 'number of entities' or indeed in some cases the 'number of events'.

To generalise the measurement of numerosity 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. it 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:

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

or a numerosity measurement would take the shape of:

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

for example:

$$\text{Population of Seborga} = 297 \text{ p}ersons \qquad (8)$$

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 us 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 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 constructing via addition 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.

The approach presented here for the numerosity of sets deviates from the baseline recommendation formulated in the SI-system and the standard frameworks presented in 2.1, nonetheless it is not unprecedented. Even these recommendations tend to acknowledge the need for measuring the 'number of entities', occasionally make way for exceptions, instead of simply calling quantities unitless. E.g. UCUM introduces specific units of measure that derive from the unit '1', e.g. bit or byte (8 bits). Likewise, in various groups of data practitioners we find cases of systematic uses of counting units (*person* and *enterprise* in economic and social statistics, *pixel*, *frame* in image processing, *revolutions per minute* in physics and engineering, packaging units *packet*, *roll* in UN/CEFACT recommendations, etc.) For further justification and a thorough description of the proposed approach (with the exception of monetary value and currencies) please refer to the David Flater papers [DF1] [DF2].

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, however as much as possible stay balanced, without and extreme bias towards precision. More precision is great, and it is easier for the end-user to generalise the scope when needed, but it may lead to a proliferation of units of measures and can lead to a scattering or isolation of the data.

4. Do not worry about the potential redundancy that might emerge between the *measure* and *unit of measure* wording - this may occur because the set to be measured may be defined by its elements, or with a reference to its typical elements.

2.6 Measuring monetary value

The concept of a *monetary value* can be formally defined as a distinct quantity kind, and indeed a new dimension. While this quantity is not one of the base quantities in the SI system, it is essential for fields like economics and social sciences. In association with *monetary value*, currencies such as the Euro or US Dollar function as units of measure, like how the metre is a unit for length. Just like for traditional physical units, monetary value can also be expressed in terms of other reference quantities representing a monetary value, such as the value of a certain amount of gold, the economic output of an entire country (GDP), or the valuation of a company.

Exchange Rates are a specific type of derived quantity that represents the ratio of two different currency units. An exchange rate, such as the US Dollar to Euro rate (USD/EUR), allows for the conversion of a monetary value from one currency to another. It can be seen as either a ratio of two currency units or as a single currency unit expressed in terms of another.

The peculiarity of *currencies* as units of measure and *exchange rates* as conversion factors among those, is that they are not fixed in time (unlike units of measures within the SI-system). The variability in time for the currencies is due to inflation (as if we defined a rubbery metre rod, which compresses or stretches in time). As inflation and other economic factors affect different currency zones at different rates, the exchange rates themselves change dynamically. Recommendations specific to modelling the inflation effects, or various conversion types are discussed at more length in section 3.2 under the heading *Price base and Conversion type*.

A key distinction in monetary value measurement is the difference between *stocks* and *flows*. A *monetary stock* refers to a value at a specific point in time, such as the total amount of money in a bank account on a given day. This concept is analogous to "voltage" in an electrical circuit, which is a potential at a single point. In contrast, a *monetary flow* represents the movement of value over a period, such as a monthly income or a nation's trade balance. This is like "current" in electricity, which measures the rate of flow of charge. Understanding this distinction is crucial for accurately describing and analysing economic phenomena, whether they are static values or dynamic

processes. Nonetheless, by convention this distinction is often omitted in databases containing economic data.

GDP, expressed in USD or national currency, carries a 'wrong' unit of measure in most statistical databases. This is because GDP being a measure of flow, i.e. value added generated over a period, 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[3], 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. (Because of the annualization convention, it is immediately comparable.) Such mental gymnastics could be spared if the unit of measure for debt-to-GDP ratio would be *year* or *quarter* as it follows from dimensional analysis.

Many essential economic quantities are derived by combining monetary value with other physical or abstract quantities. *Prices* are a prime example; they represent the ratio of a monetary value to a volume, mass, or some other quantity. For instance, the price of gasoline might be expressed as US dollars per litre, or the price of a precious metal might be given in Euros per gram. This type of derived quantity allows for the standardization of value across different items, making it possible to compare the cost of goods and services on a per-unit basis. Another example of a derived quantity is *Cost-Efficiency*, which measures the output or performance of a system relative to the monetary value invested. This can be expressed in many ways, such as miles per US dollar for transportation or a specific performance metric per monetary unit for a business process.

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

In a data-warehousing context data is structured in multidimensional cubes[4]. SDMX (statistical data and metadata exchange) standard provides a rich language to describe such cubes. [5] In this section

---

[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.

[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

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 $Measure = \{x\}$[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):

$$(Dimension_1, \ldots, Dimension_n) = \{x\}(Attribute_1, \ldots, Attribute_k) \qquad (8)$$

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 one would choose a component to be a dimension or an attribute in a Data Structure Definition in general.

Based on the SDMX guidelines [DSD] [DM], 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

---

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'.)

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).

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 – an 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		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	A multiplier to calculate the actual value in the narrow unit of measure.	Units, Thousands, Millions
TRANSFORMATION	Transformation	SDMX:CL_TIMETRANS	Time-related operation performed on a time series, involving only observations of that time series.	Index, Growth rate
ADJUSTMENT	Adjustment	SDMX:CL_SEASONAL_ADJUST	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 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 code-list) 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 quantity – 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 varying 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 seasonally adjusted variant of a quantity 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 has potentially a different value in another period – 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, or 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 [G], 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 cannot 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 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 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, an annotation should be added to reference the official UCUM, the SI-reference point and many of the popular unit of measure ontologies. When referring to kilograms, the code is set to "KG" in the proposed Units of Measure code-list. Since the official UCUM representation is "kg", the following annotation would decorate 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 and labelling

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

Itemizable 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
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 because of combining currency and scientific/itemizable 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/itemizable/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, nonetheless a good balance should be found between precision and contextuality. When shares of a measure are presented along breakdowns, there are three possible approaches: present for each breakdown a specific unit of measure (e.g. Percentage of women aged 15-49 years), or present the unit of measure with a contextual reference (e.g. Percentage of women in the same age-group), or indeed omit entirely the breakdown reference when it is obvious from the context (e.g. Percentage of women). The first option is the best at unlocking the full power of dimensional analysis, but it has the risk of scattering the data in a database (in the sense that few numbers will be immediately comparable), the second option is best when there is ambiguity about the breakdown to be considered as relevant for the unit of measure, whereas the third - least specific - option works best when it is clear from the context or the data model how 'Percentage of women' can be interpreted.

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-group
PT_POP_SEX_AGE: Percentage of population in the same sex and age-group
PT_POP_SUB: Percentage of population in the same subgroup

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 valuable 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 unit 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" 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 *itemizable*. 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 colour, 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 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 further additive or multiplicative extensions of the measured quantities. Presenting scale information for such measurements may be more relevant than describing a *unit of measure*. 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 code-list 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 require additions to reflect particularities in statistical data products. Examples of dynamic code-lists include derived units of measure, percentages, and factors. 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.

4. References

[SI] The international System of Units, 9th edition [Brochure 9](#)

[ISO] ISO 80000-1:2022, Quantities and units, [ISO](#)

[UC] The Unified Code for Units of Measure, [UCUM](#)

[UNC] UN/CEFACT code-list recommendations, Rec.20, Codes for Units of Measure in International Trade https://unece.org/code-list-recommendations

[KS] Jan Martin Keil, Sirko Schindler, Comparison and Evaluation of Ontologies for Units of Measurement, in [Semantic Web Journal 2018](#).

[DF1] Flater, D. (2023), Dealing with counts and other quantal quantities in quantity calculus, [Measurement: Journal of the International Measurement Confederation (IMEKO)](#)

[DF2] Flater, D. (2024), Letter to the editor: Unit one is intrusive, [Metrologia](#)

[DSD] [Guidelines for SDMX Data Structure Definitions](#)

[DM] [Modelling Statistical Domains in SDMX](#)

[G] [SDMX Glossary](#)