Working Paper

INTERNATIONAL MONETARY FUND

WP/11/152

IMF Working Paper

Towards a Best Practice of Modeling Unit of Measure and Related Statistical Metadata

Michaela Denk, Wilfried Grossmann

INTERNATIONAL MONETARY FUND

 WP/11/152

IMF Working Paper

STA

Towards a Best Practice of Modeling Unit of Measure and Related Statistical Metadata

Prepared by Michaela Denk, Wilfried Grossmann

Authorized for distribution by Ann McPhail

June 2011

Abstract

Data and metadata exchange between organizations requires a common language for describing structure and content of statistical data and metadata. The SDMX consortium develops content oriented guidelines (COG) recommending harmonized cross-domain concepts and terminology to increase the efficiency of (meta-) data exchange. A recent challenge is a recommended code list for the unit of measure. Based on examples from SDMX sponsor organizations this paper analyses the diversity of "unit of measure" as used in practice, including potential breakdowns and interdependencies of the respective meta-information that go beyond the current SDMX COG recommendations and possible value domains for the identified components.

JEL Classification Numbers: C49, C80, C82, C89

Keywords: SDMX, metadata harmonization, data dissemination and exchange, data warehouse

Acknowledgements: Special thanks go to our colleagues at the IMF Statistics Department, especially Mike Seiferling and Gareth McGuinness, and Karl A. Froeschl from the University of Vienna for valuable feedback and discussions on this and previous versions of the paper.

Authors‟ E-Mail Addresses: mdenk@imf.org, wilfried.grossmann@univie.ac.at

Contents Page

Tables

I. INTRODUCTION

Data and metadata exchange between statistical organizations requires a common language for describing structure and content of statistical data and metadata. In this context, the Statistical Data and Metadata Exchange (SDMX) consortium develops technical exchange standards as well as content oriented guidelines that recommend harmonized cross-domain concepts and terminology to increase the efficiency of data and metadata exchange. One of the major current challenges of this harmonization effort is a recommended code list for the unit of measure.[1]

An essential prerequisite for the harmonization of units of measure is the identification of its basic building blocks and their interrelations. In practice, diverse meta-information is packed into what is broadly referred to as "unit of measure", for instance comparison periods of growth rates, numerator / denominator of ratios, index types, scaling factors, and aggregation functions, just to name a few. The current SDMX content-oriented guidelines (COG) (SDMX Initiative, 2009) treat only some of these components separately.

Based on examples of current practices from a wide range of international organizations, this paper thoroughly analyses the diversity of "unit of measure" in practice, including potential further decomposition and interdependencies of the respective meta-information beyond the current SDMX recommendations as well as possible value domains for the identified components.

The remainder of the paper is organized as follows. Section 2 investigates international organizations' current practice of representing unit of measure. Section 3 introduces various existing standards and guidelines for unit of measure, in particular the relevant parts of the SDMX COG (Annex 1 & 2). Section 4 discusses different modeling requirements for different stages in the life cycle of statistical products, develops an approach to modeling unit of measure and related concepts for data and metadata exchange that extends the suggestions of the current SDMX COG, and proposes the usage of available standards where appropriate. As the SDMX COG are continuously evolving, this can be regarded as a suggestion for the further development of the guidelines. Section 5 provides a short wrap-up and an outlook on potential future developments.

II. CURRENT PRACTICE

In order to illustrate the current practice of modeling the concept of unit of measure and related concepts, this paper shows examples of the usage and representation of the unit of measure in dissemination databases of various international organizations as well as in IMF

[1] Other code lists, for example for the reference area (most often a country or region in international statistics), are also under development.

internal databases. These examples also serve as the basis for recommendations on further developments of the SDMX Content Oriented Guidelines as well as on the representation of unit of measure and related concepts in different stages of the life cycle of statistical data products. Examples from the following databases are used (web addresses are provided in the References section):

1. Principal Global Indicators (PGI) Website of the Inter-Agency Group on Economic and Financial Data (IAG)

2. World Economic Outlook (WEO) of the International Monetary Fund (IMF)

3. OECD.Stat Data Warehouse of the Organization for Economic Cooperation and Development (OECD), three datasets

4. Statistics Website of the Bank for International Settlements (BIS)

5. Statistical Data Warehouse of the European Central Bank (ECB)

6. Eurostat's Selected Principal European Economic Indicators (PEEI)

7. Eurostat's EUROIND Database

8. World Development Indicators (WDI) Online Database (World Bank)

9. Millennium Development Goals Indicators Database (UNStats)

10. IMF Internal Databases, three datasets

Of these 14 datasets, four do not separate the economic indicator from the unit of measure or provide the unit information as reference metadata, whereas four datasets split concepts such as unit multiplier or adjustment method from the unit. The remaining six databases separate unit of measure from economic indicator. These differences are characterized in six tables shown in the appendix providing a selection of the units used in the surveyed databases. Table 3 shows examples for a mixed dimension that combines at least information on measured (economic) indicator, type of unit, unit of measure, adjustment method, and frequency. It should be noted that several items (e.g. “Personal computers” or “Youth unemployment rate, aged 15-24, men”) fail to provide the unit information completely, assuming that it is obvious from the indicator used. Table 3 also illustrates the variety of types of units used in official statistics such as index, count, ratio, rate, percentage, or changes.

Table 4 lists 20 examples with the unit presented as a separate dimension. Along with the information contained in Table 3 it is clear that unit descriptors for the same underlying conceptual unit are heterogeneous across databases, e.g. “Percent Change over Previous

Period", "Growth previous period", and "Percentage Change", or "%", "Percent", and "Percentage Rate". Besides, some databases are less generic in dealing with changes and include the frequency in the unit, e.g. "% (Q/Q-1)" that denotes "Percent Change over Previous Period" for quarters. With respect to the decomposition of units, some of these databases make use of a few very generic units, or types of units (i.e. "Percent", "Level, ratio, or USD Millions", or "Index"), whereas other databases pack a lot of information into this concept. Some examples of these are "Per head, at current prices and current PPPs, in US dollars", and "In percentage of Net National Income". The latter is also an example for an economic indicator that is defined as the ratio of two other economic indicators but for which the denominator of the ratio ("Net National Income") is included in the unit instead of (or in addition to) the indicator itself. Moreover, in various situations "Index" can be seen without any further specification, with base period, with base period and adjustment method, and even with a "base method" stating the method used to calculate the value of the time series over the specified period that is set to the base value (e.g. moving average).

Table 5 shows examples with separate unit and unit multiplier. The latter appears under different names across databases. Some examples of these are magnitude, power code, power of ten, (unit) multiplier, or scale, and the values are either specified as multiples of 10 (e.g. 0.1), as the corresponding exponents (e.g. -1), or as text (tenth). Also, some databases set the multiplier missing instead of 0 (for 10^0), 1, or "units" in case of plain values to be multiplied by 1. In comparison to the previous table, a certain ambiguity appears. While some of these databases include concepts or distinctions, such as constant vs. current prices, the denominator of a ratio (e.g. "of potential GDP", "per capita"), or changes, in the economic indicator, other databases put this information into the unit. Equivalent information is provided in all investigated databases, but the first group decomposes the information and (at least) partially identifies relevant components, whereas the second group subsumes the entire information under a single dimension. There does not seem to be general agreement on the most appropriate level of information disaggregation. This uncertainty about where to include a particular concept can be considered a reasonable argument in favor of treating this concept separately.

Tables 6, 7, and 8 show examples for separating further concepts from indicator and unit, viz. type of measure and types of adjustment. In Table 8, the unit of measure is not represented as a dimension with a definite (coded) value domain. Instead, the respective information is included in a free text field that contains various referential metadata items. The examples in Table 6 contain a mixed unit and multiplier dimension but a separate type of measure. The type of measure discerns between "change" and what is called "level" in some of the examples in Table 4. Adjustment information is included in the type of measure as well. The examples in Table 7 also combine unit and multiplier in one dimension. Apart from that, two different types of adjustment are presented separately, viz. seasonal and price adjustment. This makes sense, as all four combinations of the values of the two types of adjustment are observable. In other words, both, current price data and chain-linked data, may be seasonally

adjusted or not. The examples in Table 8 contain type of unit (index), base value, base period, underlying currency, and index unit combined in a free text field. The type of measure is presented separately.

III. Standards and Guidelines

The existing diversity of implementation and representation of the unit of measure and related concepts in official statistics' practice may suggest a lack of standards and guidelines or at least of their adoption. This is not to say that such standards and guidelines do not exist, but not all originate from the official statistics community. The SDMX COG is regarded as the most prominent effort from within (the community) focusing on the harmonization of cross-domain concepts for (meta-) data exchange. The standardization of units of measure from a natural scientific perspective is the aim of the International System of Units and the Unified Code for Units of Measure. These units, though mainly applicable to the measurement of physical quantities like mass or length, are partly relevant in official statistics as well. Moreover, the broad structure of these systems consisting of, (i) base units, (ii) derived units (defined as being based on relations between quantities), and (iii) dimensionless units, are generalizable in the context of official statistics and therefore likely worth adopting.

A. SDMX Content Oriented Guidelines (COG)

SDMX is an initiative to foster standards for the exchange of statistical information sponsored by BIS, ECB, EUROSTAT, IMF, OECD, UN, and World Bank. The focus is on business practices in the field of statistical information that increase the efficiency of data and metadata exchange. A precondition for the automated production, processing, and exchange of data and metadata between national and international statistical organizations is the standardization of file formats for data and metadata as well as the contents of these files. In recognition of this, the sponsoring institutions have developed common technical standards and content-oriented guidelines. The SDMX COG recommend practices for creating interoperable data and metadata sets using the SDMX technical standards with the intent of generic applicability across statistical subject-matter domains, building on the cross-domain harmonization of concepts.

The SDMX COG cover a broad range of cross-domain concepts such as sampling, data collection, methodology, data quality, administrative information, reference area, and time. Four of the COG cross-domain concepts are related to the unit of measure, viz. unit of measure (item 65), adjustment (3), base period (5), and unit multiplier (64). The representation of unit of measure, adjustment, and base period is divided into a code and free text, but no code lists are provided. For base period, a time stamp is also required. Unit multiplier is the only related concept with a code list available.

Unit of measure is defined by three components, (i) type (currency is named as an example), (ii) unit code, and (iii) unit of measure detail. „Type" is not explained or used any further in the SDMX COG. For units that are index numbers, index type exists as a separate cross-domain concept. „Unit" is defined as "a quantity or increment by which something is counted or described"; some examples are listed such as kg, Euro, counts, and index numbers.

Unit multiplier specifies the exponent to the basis 10 by which observation values were divided, usually for presentation purposes. For instance, a unit multiplier of 3 indicates that the data are provided in thousands. The code list does not contain negative unit multipliers. It is referred to as a supplementary concept necessary to interpret observed values. Unit of measure detail is intended to hold any additional information required to understand the unit in detail, and no connection is made to other related concepts such as adjustment and base period.

Base period is relevant for the interpretation of index data, series at real terms, e.g. data at constant prices, change measures such as percentage changes with respect to the previous period, or other series based on a certain point in time (in addition to the reference period). The base period of an index is the period when the index equals the base value (per definition). This period may range from a single day, for instance for stock exchange indices, to a couple of years or even longer time intervals. Although the value at the base period is 100 in many cases, this base value needs to be specified as well. In practice, the base period is included in the specification of the unit of measure or the indicator quite frequently, and the base value is often omitted and assumed to equal 100. The SDMX COG contain base period as a concept, but do not suggest a separate concept for base value. They merely indicate that the base period may include information on the base value, most likely in the free text part. As the code list is not (yet) available, its intent is not completely clear. It seems reasonable to assume that it should distinguish between different types of base periods such as month, year ranges, or relative base periods (e.g. "previous period").

Adjustment is a concept that is included in the unit of measure or the economic indicator in many statistical databases, as illustrated in section II and the appendices. The SDMX COG treat this concept separately and define it as being associated with changes in definitions, exchange rates, prices, seasons, and other factors. For means of presentation, the concept is split into code (the type of adjustment) and free text. Instead of a code list for adjustment only some examples referring to time series adjustment such as trading day, seasonal, or trend-cycle, are provided. The free text is intended to contain any additional detail on the adjustment method.

B. International System of Units (SI)

The International System of Units (Bureau International des Poids et Mesures, 2010) was established and is defined by the General Conference on Weights and Measures. Its origin can be traced back to the creation of the decimal metric system at the time of the French Revolution and the subsequent deposition of two platinum standards representing meter and

kilogram in the Archives de la République in Paris in 1799. It is an evolving system of units of physical measurement based on a system of quantities and equations defining the relations among the quantities. Most of the quantities are listed in the International Standard for Quantities and Units ISO 80000 (2009).

The foundation of the system are the seven base quantities length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity that are by convention assumed to be independent. The corresponding base units are meter, kilogram, second, ampere, Kelvin, mole, and candela. Derived units are created by combining base units according to the algebraic relations linking the corresponding quantities. Currently there are 22 SI derived units. In addition, the SI permits the use of certain supplemental units such as the traditional units of civil time (minute, hour, day, and year). For scaling, i.e. creating multiples of approved units, such as km as 10^3m, the SI specifies a list of prefixes ranging from yotta- at 10^{24} to yocto- at 10^{-24}.

C. Unified Code for Units of Measure (UCUM)

The Unified Code for Units of Measure (Schadow, McDonald, 2009) is a code system intended to include all units of measure contemporarily used in international science, engineering, and business. The purpose is to facilitate unambiguous electronic communication of quantities together with their units. It is based on ISO and ANSI standards but resolves numerous name conflicts present and has a broader coverage by including units that are not covered by ISO standards although relevant in practice. Similar to the SI, seven base units (partly differing from the SI base units) and their respective base quantities are distinguished, supplemented by derived units, units from other standards, international customary units, and US or British Imperial units. In addition, the UCUM defines "dimensionless" units used for simple counts or quantities derived as a ratio from quantities measured in the same unit that are often provided in percent.

IV. Beyond the Current SDMX COG

Comparing the 14 examples from official statistics with the available standards and guidelines indicates that the existing standards and guidelines (i) are only partially met in practice and (ii) incompletely cover or specify the concepts required. Although the two unit standards used for physical quantities do not cover all units relevant for measuring social and economic phenomena, it makes sense to use them for the generalizable domains and to adopt the principles they follow to define units.

None of the 14 investigated databases entirely follow the recommendations of the SDMX COG in terms of the separation of unit and measurement information from the economic indicator measured and its decomposition into unit, unit multiplier, base period, and adjustment. Four databases do not split economic indicator and unit of measure or just use reference metadata to provide the unit information. Six databases separate the unit from the

indicator, but are heterogeneous with respect to the presentation of base period, adjustment, and unit multiplier. Some include base period and adjustment in the indicator, some in the unit. The unit multiplier is part of the unit in most cases. The remaining four datasets split one or more additional concepts from indicator and unit, but they differ regarding what concept is represented separately. One of the databases has a separate unit multiplier. Two databases use a 'type of measure' that contains information on the adjustment for one database and on aggregation methodology for the other database. Finally, one database splits two different types of adjustment from unit and indicator, but fails to provide unit multiplier and base period separately.

One may certainly argue that the majority of the 14 examples presented above stem from online dissemination databases that need not necessarily follow guidelines recommended for inter-organizational data exchange, which may serve as an explanation for the deviations from the SDMX COG. However, even in data dissemination that does not primarily aim at the systematic, electronic consumption of the published data, a certain level of structural and semantic harmonization seems reasonable to facilitate comparison of data. Data modeling at different stages of the life cycle of statistical products such as data collection, processing, publication, and exchange differ in requirements and priorities. For example, in data collection, listing only the admissible value combinations of several dimensions in one "mixed" dimension easily prevents users from providing data for invalid combinations of dimension values as would be possible in a database that presents the dimensions separately. This is to say that in the latter situation constraints on the precluded cells (also called "structural zeros") have to be defined and checked, whereas in the former situation the invalid value combinations are just not included in the value domain of the mixed dimension. Another reason for considering mixed dimensions as advantageous for publication or particular collection purposes is their better comprehensibility for users. Still, this is not a generalizable fact, but rather what can be assumed for more executive types of users such as decision makers or journalists. Users that intend to work with the data as for instance students or researchers may need to merge data from several sources for their analysis. These more analytic types of users would also benefit from a finer modeling granularity. In any case, the finest modeling granularity is preferable for data and metadata exchange between organizations. Splitting mixed dimensions into their components creates a purer, but higher-dimensional model that allows data recipients to assemble these components to conform to their internal data models.

Despite the imperfect compliance to the SDMX COG in official statistics dissemination practice, these guidelines still do not seem sufficiently specific and restrictive to account for the requirements derived from the investigated examples, especially concerning the decomposition of unit of measure and related concepts as well as the guidance on value domains (code lists) for the resulting components.

In order to derive detailed suggestions in this respect we revisit the examples of section 2 by decomposing units into their necessary components and developing basic, extendable value domains for these components as a foundation for the specification of suitable code lists.

A. Generic Unit Types

The units from the collected examples can be classified into a few generic unit types that also reflect the distinction between actual measurement (and units for measured quantities) and calculations (and units for calculated quantities):

Base types can be broken down into the categories of count, currency, and (a subset of) the base quantities of SI and/or UCUM (e.g. length, mass, or time).[2]

For unit type **count** the unit itself specifies what is actually counted (e.g. persons, personal computers, mobile internet subscriptions, jobs). In addition, a unit multiplier and the method of time series adjustment are required.

For unit type **currency** the unit provides the actual currency or currency group, for example Euro, SDR (special drawing rights), or domestic currency. Again, unit multiplier and method of time series adjustment must also be specified. In addition, currency units require information on the method of price adjustment, such as constant prices (real values), current prices (nominal values), or chain-linked, and method of exchange rate adjustment, e.g. constant. Irrespective of the adjustment type (time series, price, etc.) all "constant" methods of adjustment as well as adjustments making use of an index (e.g. chain-linked) require the specification of a base period.

Derived unit types can be defined by the derivation method applied, i.e. the type of calculation carried out on base (or derived) unit types to obtain a new unit type, such as difference, sum, ratio, product. These can be broken down into categories of: ratio, product, difference, and sum. The derived quantities of SI/UCUM such as area, volume, and velocity can be regarded as subtypes thereof. The list of derived types is extendable by any arithmetic operator (or other calculation) that can be applied to base (and derived) unit types and units to create new derived types and units.

Frequently, derived units (and unit types) appear in case of indicators derived from other indicators, for example "Unemployment Rate" as ratio of the number of unemployed persons in the labor force to the total labor force, "Foreign Debt in % of GDP", "GDP per capita" as ratio of GDP and population, or "Quarterly Absolute Changes in External Positions of

[2] Examples for physical quantity types such as length, mass (base), or area (derived as product of two length types) are not discussed any further in this paper. They can be handled by referring to SI/UCUM. The only additional dimension that needs to be specified for these unit types is the unit multiplier.

Banks" as difference between the external positions of banks at the end of the current and of the previous quarter. In some cases, the original indicators from which a new one is derived may not be that obvious, e.g. for "Real Wages in USD/hour" calculated as ratio of total real wages to total hours worked. Moreover, derived unit types and units also occur for indicators and quantities that one would usually not consider as derived, although technically they are. Examples include exchange rates (e.g. "Currency exchange rates as national units per US dollar"), interest rates (typically specified as percent per annum), derived physical units (e.g. square meter for area, Kbit/second and km/h for different types of speed), or prices per unit (e.g. "Oil price in US dollar per barrel").

This implies that derived indicators and units can be specified by referencing the underlying indicators and/or units (which may themselves be base or derived). The unit type, unit, and multiplier (codes and descriptors) can be "calculated" from the corresponding information attached to the original indicators and/or units, such as USD/hour, Euro/USD, national currency/person. Ratios of (indicators measured in) the same unit such as external debt/GDP or unemployment rate actually result in a dimensionless quantity (USD/USD=1, person/person=1) which may be indicated by a special unit. Differences are only admissible for identical units. In addition to the derived codes and descriptors, it may be helpful to assign less technical descriptors for final data consumers.

Overall, this recursive approach allows great flexibility in defining a broad variety of derived units on demand without the need of enumerating all possible combinations in advance. Nevertheless, it seems helpful to pre-specify a broad family of derived units and attach familiar descriptors to enhance ease of use and comparability. Another option may be the definition of more specific derived unit types as children of the generic types, not only to establish frequently used groups of units but also because of similar structural characteristics of units belonging to the same subtype. Examples for possible unit subtypes are **area** and **volume** as children of product, **absolute change over reference period** as child of difference, and **currency exchange rate**, **relative change over reference period**, **index**, and **velocity** as children of ratio. The definition of a unit subtype should include the specification of the underlying unit types (e.g. unit of numerator = length and unit of denominator = time for unit subtype velocity). In addition, restrictions on units, multipliers, or even other related concepts may be necessary.

The requirement to specify a base period (in absolute or relative terms) for units of change subtypes may serve as an example for the shared structural characteristics of units of the same subtype. An absolute base period is specified as a date (part), e.g. year or quarter, whereas a relative base period is defined relative to the time period (and subject to the frequency) of the observation, such as previous period, same period of previous year, or last day of previous month. Table 10 in the appendix shows selected examples from section 2 rearranged by derived unit type.

In official statistics, ratios are of particular relevance as they represent a basic principle of specifying comparable social and economic indicators. Relative changes of indicators are also frequently used in official statistics, also for reasons of comparability between countries and over time. The following table provides an overview of different cases to be distinguished for unit type ratio depending on the comparison of unit type, unit, and unit multiplier of the underlying numerator and denominator unit types, units, and multipliers.

The unit type in this case is always ratio, although subtypes that are more specific may be used. Therefore, this column is omitted from the table. The possible subtypes mentioned are oriented at the examples; another feasible approach is to base subtypes on the eight distinct cases according to the comparison of numerator and denominator type, unit, and multiplier.

For subtype **index** the detailed specification contains base period, base value, and methods of different types of adjustment. The unit multiplier is usually 1. For stock indices like the last two examples in Table 3 the aggregation method, e.g. end of period (last), moving average (plus reference period), is relevant as well. In addition, further methodological metadata may be provided such as the type (e.g. Laspeyres) or weighting system as defined in Annex 4 in the SDMX COG.

The units and unit multipliers are calculated as ratios of the units and multipliers of the numerators and denominators. Ratios of the same unit and unit type in the numerator and denominator result in a dimensionless quantity as cancelling the ratio of the units yields 1. We term the resulting unit simply as "fraction". In cases where the units in numerator and denominator differ, the unit of the result is the ratio of the original units and may be labeled "U_N per U_D". If the unit multipliers coincide, the multiplier of the ratio equals 1. However, the values in the unit multiplier column of the result are only valid if the ratio is not rescaled. For example, "External debt in Euro, Millions" as ratio of "GDP in Euro, Millions" is only measured in units (multiplier=$10^6/10^6$=1) unless it is rescaled to percent (multiplier=0.01).

This also shows that, strictly speaking, "percent" is not a unit, but rather a unit multiplier, especially relevant for dimensionless quantities. For fractions, the most important additional information is the unit multiplier. In addition, the underlying unit ("base unit") should be provided (e.g. Euro for the "External debt as % of GDP" example).

Where the unit type is not the same for numerator and denominator, one may regard it as an option to set the resulting unit type to $Type_N/Type_D$ (as a subtype of ratio). At a first glance, the cases with different unit type but same unit may seem unfeasible. Still, they apply at least to relative changes over a reference period. As stated above, relative changes are derived from a difference (absolute change) and the unit type underlying this difference, but even require the units to be the same in the numerator and denominator. The multipliers may differ.

Table 1. Unit Type Ration: Case Distinction

Numerator & Denominator			Result (Ratio)		Possible Subtypes & Examples
Type[1]	Unit[1]	Multiplier[1]	Unit	Multiplier	
=	=	=	fraction $U_N/U_D=1$	$M_N/M_D=1$	**Rate/Share of total**, e.g. Share of youth unemployed to total unemployed; Military spending as % of GDP **Ratio of subgroups**, e.g. Women to men parity index, as ratio of literacy rates, 15-24 years old **Ratios between same currency units**, e.g. External Debt as % of GDP **Index**, e.g. Inflation, average consumer prices, Index, 2000=100
=	=	different	fraction	M_N/M_D	**Rate of total (in 10x)**, e.g. Malaria death rate per 100,000 population
=	diff.	=	U_N/U_D	1	**Exchange rates**, e.g. Purchasing power parities conversion factor, local currency unit to international dollar
=	diff.	different	U_N/U_D	M_N/M_D	**(Interest) Rate** in % per annum, e.g. Overnight interbank rate, Percent per annum
different	diff.	=	U_N/U_D	1	**Ratios of physical units**, e.g. km/h Ratios of currency per physical unit or count, e.g. USD/barrel; USD/hour; or Euro/capita
different	diff.	different	U_N/U_D	M_N/M_D	e.g. Energy use (kg oil equivalent) per $1,000 GDP – kg/1000 USD
different	=	=	fraction	1	**Relative Change over Reference Period**, e.g. GDP % Change over previous period
different	=	different	fraction	M_N/M_D	Relative Change over Reference Period, with **different unit multipliers** used for numerator and denominator

[1] An = indicates that the component (type, unit, or multiplier) is the same for numerator and denominator. U is short for unit, M for Multiplier. The subscripts N and D stand for numerator and denominator, respectively. The table points up the complexity and extent of the analysis if conducted at a certain level of detail.

Two special cases related to percentages are "Interest rate in % per annum" and the difference of ratios. For the former, the underlying unit types and units are ratio (or rate/share of total) and fraction in the numerator and time and year in the denominator. For the latter assume that both ratios are provided as fraction with multiplier 0.01 (=%). Then the result is actually measured in "percentage points". This implies that it may be reasonable to assign that as a common descriptor to the combination of the concepts unit type (=difference), unit (=fraction), and multiplier (=0.01). This idea of a joint descriptor for a combination of particular values of a set of concepts related to unit of measure does not prove the efforts of decomposing unit of measure into its components wrong. Rather, it serves as evidence that what makes sense for data users may be quite different from what is reasonable for electronic data processing and exchange.

Another special ratio subtype is the ratio of subgroups typically used for balance indicators such as gender parity indices (e.g. "Women to men parity index, as ratio of literacy rates") or economic balance indicators. For this type of indicators, a "base value" indicates the level that corresponds to a balanced situation (gender parity in the first example); values below or

above this threshold indicate (negative or positive) imbalance. This base value is usually 1 or 100 but should be provided as additional information, as the data may be rescaled to a different imbalance threshold.

B. Unit of Measure and Related Concepts

Recapitulating the observations made from the examples and the exercise of specifying a set of generic unit types and subtypes, we obtain the following decomposition of unit of measure.

1. **Unit Type** as discussed above

2. **Unit** depending on Unit Type (see table below)

3. **Unit Multiplier** with some restrictions based on Unit Type and Unit, for example Units with Unit Type=Index usually have Multiplier=1

4. **Base Period** with value domain including time stamps (in different granularity) extended by predefined values such as previous period, corresponding period of previous year, previous three months, December of previous year, years since time stamp, moving average of years since time stamp (moving average and similar methods that define how the value is calculated in the base period could even be put into a separate concept (or sub-concept).
   Mandatory for: Unit Type=Index, Absolute or Relative Change over Base Period; Price Adjustment=Constant Prices, Reference Chained, Chain-Linked

5. **Base Value** with default value 100
   Mandatory for: Unit Type=Index or Balance Indicator

6. **Adjustment** split into (at least) three components

   a. Price Adjustment such as Constant, Current, Reference Chained, Chain-Linked
      Mandatory for: Unit Type=Currency, Derived and Base Unit=Currency (recursion)

   b. Econometric/Time Series Adjustment such as Seasonal, Working Day, Trading Day, Trend Cycle, Amplitude, Ratio to Trend MCD Smoothed

   c. Exchange Rate Adjustment such as Constant, Current
      Mandatory for: Unit Type=Currency, Derived and Base Unit=Currency (recursion)

7. **Base Unit** refers to the underlying Unit of the components of derived Units, especially of a Difference (minuend and subtrahend) or of a Fraction (numerator and denominator)
Mandatory for: Unit Type=Difference; Unit=Fraction

8. **Base Indicator** refers to the components of a derived Indicator that may be the reason for the usage of a derived Unit; role of the Indicator must be specified (e.g. numerator)
Mandatory for: Unit Type=Ratio (with a few exceptions, such as SI/UCUM Unit Types)

Table 2. Unit Types, Units, and Constraints

Unit Type	Unit	Constraints
Base		
Count	Person	-
	Mobile Phone	-
	…	...
Currency	List according to ISO 4217 (2010)	Requires Price Adjustment
	Different currency groups	Requires Price Adjustment
	National Currency	Requires Price Adjustment
Length	List according to SI/UCUM	-
Other relevant SI/UCUM base quantities	Lists according to SI/UCUM	-
Derived		
Difference	= Base Unit	Constraints on admissible Units may apply
Absolute Change over Reference Period	= Base Unit	Requires Price Adjustment if Type of Base Unit = Currency
Balance Indicator Diff.	= Base Unit	requires Base Value
Other relevant subtypes	...	...
Product	= Product of Base Units	Constraints on admissible Units may apply
Area	List according to SI/UCUM	-
Volume	List according to SI/UCUM	-
...	...	...
Ratio	= Ratio of Base Units if Ratio=1 then Unit=Fraction	Constraints on admissible Units may apply
Velocity	List according to SI/UCUM	-
...	...	...
Numerator=Denominator	Fraction	Base Unit and Type of Base Unit the same for denominator and numerator
Balance Indicator Ratio	Fraction	Base Unit=Fraction, requires Base Value
Index	Fraction	Requires Base Period, Base Value Unit Multiplier = 1
Currency Exchange Rate	= Ratio of Base Units	Type of Base Units = Currency, $C_N \neq C_D$
Interest Rate	= Fraction / Year	-
Relative Change over Reference Period	= Fraction (=Difference/Type of Base Unit of Difference)	Type of Base Units must match
Other relevant subtypes	...	...
Sum	= Base Unit	Constraints on admissible Units may apply

Table 2 summarizes the suggested generic Unit Types, a couple of potential subtypes, and corresponding Units as well as constraints that apply to certain Unit Types or Units.

In addition to these concepts, **unit families**, **standard units**, and **conversion factors** (cf. Froeschl, 1997; Froeschl, Grossmann, Del Vecchio, 2003) play an important role, especially for physical (and currency) units. For example, there are metric, US, and British systems (families) of units. Within a family, one unit is designated as standard unit, and conversion factors relating each member of the unit family to that standard unit. Besides, conversion factors between the standard units of different unit families of one unit type are specified. For currencies, the time-dependency of exchange rates adds to the level of complexity. These concepts are not discussed in more detail here due to the focus on economic and social statistics and the lack of space.

V. Conclusion

This paper presents an analysis of existing guidelines and standards for and the current practice of modeling unit of measure and related concepts in official statistics, based on examples from databases of international organizations that act as sponsors of the SDMX initiative. This investigation reveals that the SDMX COG provide a reasonable basis for harmonization and standardization efforts of the unit of measure cross-domain concept, but that they face two challenges: (i) a lack of compliance in practice and (ii) room for improvement in terms of definition of concepts (including code lists) and decomposition of concepts. The degree of compliance with the SDMX COG varies between the 14 examined databases, but none of them provides the four concepts recommended in the SDMX COG, i.e. unit, unit multiplier, base period, and adjustment, separately. Moreover, there is disagreement in terms of the representation of a particular concept even between those databases that display that concept separately. One major necessity to achieve a harmonized representation is a more precise definition of the concepts and the inclusion of code lists in the SDMX COG. These findings were the motivation to further decompose the concept “unit of measure” into its basic building blocks and to derive an extendable set of generic unit types and subtypes, assigned units, and a set of cross-domain concepts related to unit of measure. The results of this conceptual disaggregation as presented in this paper provide a sound basis for an extension and further development of the SDMX COG with respect to cross-domain concepts as well as code lists in the context of unit of measure. In addition, ongoing and future work requires an extension of the presented ideas to define in greater detail the representation of derived indicators, unit types, units, and unit families as well as the required constraints. One of the overall objectives of this work is the development of a unit (type) calculus to better account for and make use of the calculability characteristic of units based on ideas developed by Froeschl (1997).

VI. References

Bank for International Settlements (2010) BIS Locational Banking Statistics. http://www.bis.org/statistics/bankstats.htm

Bureau International des Poids et Mesures (2010) The International System of Units. http://www.bipm.org/en/si/

European Central Bank (2010) ECB Data Warehouse. http://sdw.ecb.europa.eu/

Eurostat (2010) Selected Principal European Economic Indicators. http://ec.europa.eu/eurostat/euroindicators

Eurostat (2010) EUROIND Database. http://epp.eurostat.ec.europa.eu/portal/page/portal/euroindicators/database

K.A. Froeschl (1997) Metadata Management in Statistical Information Processing. Wien-New York: Springer.

K.A. Froeschl, W. Grossmann, V. Del Vecchio (2003) The Concept of Statistical Metadata. METANET Deliverable D5. 127pp.

Inter-Agency Group on Economic and Financial Statistics (2010) Principal Global Indicators Website. http://www.principalglobalindicators.org/

International Monetary Fund (2009) World Economic Outlook Database. http://www.imf.org/external/pubs/ft/weo/2009/02/weodata/index.aspx

ISO (2010) ISO 4217 Currency and Funds Name and Code Elements. http://www.iso.org/iso/support/faqs/faqs_widely_used_standards/widely_used_standards_other/currency_codes/currency_codes_list-1.htm

ISO (2009) ISO 8000 Quantities and Units. http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=30669

OECD (2010) OECD.Stat Extracts. http://stats.oecd.org/

G. Schadow, C.J. McDonald (2009) The Unified Code for Units of Measure. http://unitsofmeasure.org/ or http://aurora.regenstrief.org/~ucum/ucum.html

SDMX Initiative (2009) SDMX Content Oriented Guidelines 2009. 16pp + 5 annexes. http://www.sdmx.org/

UNStats (2010) Millennium Development Goals Indicators Database. http://mdgs.un.org/unsd/mdg/Data.aspx

World Bank (2010) World Development Indicators Online Database. http://ddp-ext.worldbank.org/ext/DDPQQ/member.do?method=getMembers&userid=1&queryId=135

APPENDIX 1. EXAMPLES BY DATA STRUCTURE

Table 3. Examples for Indicator including Unit of Measure

Share Prices, Monthly, Index (base 2005 = 100)
Industrial Production Index, Monthly, (base 2005 = 100), seasonally adjusted
Currency exchange rates, Monthly, National units per US-Dollar
Personal computers
Youth unemployment rate, aged 15-24, men
Adolescent fertility rate (births per 1,000 women ages 15-19)
Electric power consumption (kWh per capita)
Long-term interest rates, Monthly, Per cent per annum
Primary completion rate, total (% of relevant age group)
Ratio of youth unemployment rate to adult unemployment rate, men
Military expenditure (% of GDP)
GDP growth (annual %)

Table 4. Examples for separate Indicator and Unit

Economic Indicator	**Unit of Measure**
Population	Number of Persons, Millions
GDP Nominal	National Currency, Seas. Adjusted, Millions
Current account balance, USD millions, seasonally adjusted [levels]	Level, ratio, or USD Millions
Gross domestic product	Per head, current prices & current PPPs, USD
Oil Price	US Dollars per Barrel
Exchange Rates	National Curr. per US Dollar, end of period
Exchange rates, USD monthly averages [ratio]	Level, ratio, or USD Millions
Economic Sentiment Indicator	Index
GDP Volume	Index, 2005=100, Seasonally Adjusted
GDP Volume	Percent Change over Previous Period
Index of Notional Stocks, MFIs, central government and post office giro institutions report. sector - Monetary aggregate M3, All currencies combined - Euro area (changing composition) counterpart, Non-MFIs excl. central government sector, Annual growth rate, Working day & seasonally adjusted	Percentage change
Hourly earnings: manufacturing (seasonally adjusted)	Growth on same period of previous year

Table 4 continued. Examples for separate Indicator and Unit

Economic Indicator	**Unit of Measure**
Civilian employment: all persons (seasonally adjusted)	Growth previous period
GDP Nominal	% (Q/Q-1)
Leading indicator, amplitude adjusted [ratio]	Level, ratio, or USD Millions
GDP Deflator	Percent
Unemployment Rate	Percentage Rate
Interest Rates	Percent per Annum
Government expenditure	In percentage of Net National Income
Business Confidence Indicator	Long-term moving average = 100

Table 5. Examples for separate Indicator, Unit, and Unit Multiplier

Economic Indicator	**Unit of Measure**	**Multiplier**
Inflation, average consumer prices	Annual percent change	-
Inflation, average consumer prices	Index, 2000=100	-
Gross domestic product per capita, constant prices	National currency	Units
Implied PPP conversion rate	National currency per current international dollar	-
Gross domestic product based on purchasing-power-parity (PPP) share of world total	Percent	-
Six-month London interbank offered rate (LIBOR)	Percent	-
General government structural balance	Percent of potential GDP	-
Unemployment rate	Percent of total labor force	-
Employment	Persons	Millions
Current account balance	U.S. dollars	Billions

Table 6. Examples for separate Indicator, Unit, and Type of Measure

Economic Indicator	**Type of Measure**	**Unit of Measure**
External positions of banks in individual reporting countries in all currencies vis-à-vis all sectors	Amounts outstanding	Billions of US dollars
External positions of banks in individual reporting countries in all currencies vis-à-vis all sectors	Estimated exchange rate adjusted changes	Billions of US dollars

Table 7. Examples for Indicator, Unit, and Adjustments

Economic Indicator	**Unit of Measure**	**Seasonal Adjustment**	**Price Adjustment**
Taxes less subsidies on production and imports	Millions of euro (from 1.1.1999) / ECU (up to 31.12.1998)	Not seasonally adjusted data	Chain-linked volumes, reference year 2000 (at 2000 exchange rates)
GDP at market prices	Millions of euro / ECU	Seasonally adjusted & adjusted by working days	Current prices

Table 8. Examples for separate Indicator, Type of Measure, and Unit (as free text)

Economic Indicator	**Type of Measure**	**Reference Metadata (excerpt)**
NASDAQ COMPOSITE INDEX	Last price	The index was developed with a base level of 100 as of 02/05/1971. Currency: USD Index Unit: Points
MSM30 Index	Moving Average 60 Days	The index was developed with a base level of 1000 as of 06/1990. Currency: OMR Index Unit: Points

APPENDIX 2. EXAMPLES BY UNIT TYPE

Table 9. Examples for Count Type Units

Internet users		
Employment, seasonally adjusted	Persons	Millions

Table 10. Examples for Currency Type Units

Reserve Assets, Monthly, SDR millions			
Current account balance	U.S. dollars	Billions	
Taxes less subsidies on production and imports	Millions of euro (from 1.1.1999) / ECU (up to 31.12.1998)	Not seasonally adjusted data	Chain-linked volumes, reference year 2000 (at 2000 exchange rates)

Table 11. Examples for Index Type Units

Industrial Production Index, Monthly, (base 2005 = 100), s.a.		
Net barter terms of trade (2000 = 100)		
Economic Sentiment Indicator	Index	
GDP Volume	Index, 2005=100	
Inflation, average consumer prices	Index, 2000=100	-
NASDAQ COMPOSITE INDEX	Last price	The index was developed with a base level of 100 as of 02/05/1971. Currency: USD, Index Unit: Points
MSM30 Index	Moving Average 60 Days	The index was developed with a base level of 1000 as of 06/1990. Currency: OMR, Index Unit: Points

Table 12. Examples for Units of Derived Unit Type

DIFFERENCE		
Absolute Change over Reference Period		
External positions of banks in individual reporting countries in all currencies vis-à-vis all sectors	Estimated exchange rate adjusted changes	Billions of US dollars
Difference between Subgroups (Balance Indicator - Difference)		
Share of Net Income, Gender Gap	Percentage Points	
Business Confidence Indicator	Long-term moving average = 100	
RATIO		
Rate or Share of Total		
Unemployment rate	% of total labor force	-
Military expenditure (% of GDP)		
Ratio of Subgroups (Balance Indicator - Ratio)		
Women to men parity index, as ratio of literacy rates, 15-24 years old		
Rate or Share of Total (in 10^x)		
Malaria death rate per 100,000 population, ages 0-4		
Exchange Rates		
ECB reference exchange rate, US dollar/Euro, 2:15 pm C.E.T.	US dollar	
Implied PPP conversion rate	National currency per current international dollar	-
Ratio of same Currency Units		
External Debt	Percent of GDP	
Interest Rates		
Long-term interest rates, Monthly, Per cent per annum		
Interest rate, government debt	%	
Ratio of Physical Units		
Max. available broadband speed	kbit/second	
Currency per Physical Unit or Count		
Oil Price	US Dollars per Barrel	
Gross domestic product per capita, constant prices	National currency	Units
Physical Unit per Count		
Renewable internal freshwater resources per capita (cubic meters)		
CO2 emissions (metric tons per capita)		
Physical Unit per Currency (in 10^x)		
Energy use (kg oil equivalent) per $1,000 GDP (Constant 2005 PPP $)		
Relative Change over Reference Period		
GDP Volume	Percent Change over Previous Period	
Hourly earnings: manufacturing (seasonally adjusted)	Growth on the same period of the previous year	