Guidelines for

Global DSD for Labour Statistics

December 2018

# Table of Contents

[Table of Contents 2](#_Toc533149260)

[Document Version History 2](#_Toc533149261)

[Introduction 3](#_Toc533149262)

[Addendum for first public review 3](#_Toc533149263)

[Global DSD Version Notes 4](#_Toc533149264)

[Concepts and Code Lists 4](#_Toc533149265)

[The DSD Matrix workbook 9](#_Toc533149266)

[Example coding 12](#_Toc533149267)

[Availability of the SDMX Artefacts 13](#_Toc533149268)

[Governance 14](#_Toc533149269)

[Maintainence Schedule 14](#_Toc533149270)

# Document Version History

|  |  |  |
| --- | --- | --- |
| Version | Date | Remarks |
| 0.1 | 21/12/2018 | Draft for public pilot |
|  |  |  |

# Introduction

The international organisations seek to improve the reporting and exchange framework for labour statistics. The SDMX standard has been proven as a method to harmonise data exchange flows and systems, standardise IT systems, and gain efficencies and timeliness in the statistical process; therefore a project was started that would build on lessons learnt in the other Global “Data Structure Definition” (hereafter abbreviated to “DSD”) projects and the guidelines produced by the SDMX community.

The SDMX sponsor organisations (BIS, ECB, Eurostat, IMF, OECD, United Nations and World Bank) agreed that the DSD for Labour Statistics project meets the requirements of a Global DSD project. Global DSDs are seen as the “gold standard” for exchanging statistical data for the domain, and require the approval of the SDMX sponsors board. Global DSDs have strict governance, maintenance, review and design requirements, part of which are public reviews and a set of supporting documentation. The term “Global DSD” actually encapsulates all of the SDMX artefacts designed in the project (there may be multiple DSDs) and are governed by the [ownership group](#_Ownership_Group).

This guidelines document includes a description of the Labour Statistics SDMX artefacts in the model. A key part of the guidelines is the referenced “DSD Matrix” which includes the Labour Statistics concept scheme, the individual data flows, and the data flow/DSDs relationships, and codelists.

# Addendum for first public review

At the start of the public review there are some pending items which will appear shortly after the review has started:

* The dataflow constraints;
* The SDMX-ML files for the artefacts, e.g. the XML for DSDs, codelists, concept scheme, etc.
* Some indicator decompositions in the Indicator-Coding matrix are to be completed;
* The CL\_AREA group codes
* An example of qualitative data coding in the guidelines

Any changes to the ongoing public review material will be posted on SDMX.org/Public reviews.

The aim of the review is to collect feedback and raise issues and questions on any aspect of the “Global DSD for Labour statistics” material. The contact point is the project manager [david.barraclough@oecd.org](mailto:david.barraclough@oecd.org) who will maintain an issue log. The [ownership group](#_Ownership_Group) will react to the issues in the log and issue updates to the material on the public review website.

As an outcome of the public review, the final production model will incorporate changes based on the review feedback.

# Global DSD Version Notes

This chapter provides a short overview on the main changes done in the SDMX artefacts for each published version.

|  |  |  |
| --- | --- | --- |
| Version | Date | Remarks |
| 0.1 | 20/12/2018 | Public review release |
|  |  |  |

# Concepts and Code Lists

An SDMX DSD consists of several statistical concepts depending on the domain(s) being covered: e.g. institutional sectors, industries, commodities, etc. These concepts are separated into dimensions and attributes:

* Dimension concepts are used to uniquely identify a certain data item or variable and when joined together, they provide the “time series key” as the unique identifier for an observation. In an SDMX dataset, a value must be assigned to each of the dimensions relevant for a data item.
* Attribute concepts are used to further describe the data relating to, for example, observation status, confidentiality status, embargo time, base period. They can be attached at different levels of the dataset: to the whole dataset, at the series level, at the observation level, or a partial series key. Attributes are either mandatory or conditional (can be omitted in the message). Furthermore, attributes may be coded (their value is restricted by a related code list) or be uncoded, and they have a data type (e.g. string, datetime).

## Concepts in the Concept Scheme

This section describes all the concepts that are maintained in the concept scheme delivered by the SDMX Global Labour DSD project. They are listed in the *Labour Statistics DSD Matrix* workbook *Concept Scheme* worksheet.

### Dimensions

All dimensions in the Global Labour concept scheme are coded, therefore have an associated code list. See the “Concept Scheme” worksheet in the “Labour Stats DSD Matrix” workbook for exact details.

#### Measure (MEASURE)

## Data element that represents a simple aggregation and which takes one or more vectors as arguments and returns a scalar.

#### Aggregation Operation (AGGREGATION\_OPERATION)

## Representation of the calculation type when a single value is calculated from a collection of values.

#### Reference period of the indicator (INDICATOR\_REF\_PER)

The time interval to which the indicator refers.

#### Type of hours (HOURS\_TYPE)

The type of hours worked.

#### Type of value (VALUE\_TYPE)

The type of value, such as nominal or real.

#### Unit of measure (UNIT\_MEASURE)

Unit in which the data values are expressed.

#### Reference area (REF\_AREA)

Country or geographic area to which the measured statistical phenomenon relates. This concept implements the CL\_AREA national-level code list for the majority of dataflows, but for certain dataflows that require the exchange of sub-national data, the concept implements the CL\_REGIONAL code list which is based on the NUTS classification.

#### Data source (SOURCE\_TYPE)

Characteristics and components of the raw statistical data used for compiling statistical aggregates.

#### Sex (SEX)

State of being male or female.

#### Age (AGE)

Length of time that an entity has lived or existed.

#### Labour force status (LABOUR\_FORCE\_STATUS)

Classification of the working-age population based on the person's relation to the labour market, including labour market attachment.

#### Status of Worker (WORKER\_STATUS)

Refers to the nature of the relationship between a worker and the economic unit in which or for which the work is performed

#### Education level (EDUCATION\_LEV)

Based on the ISCED classification. An ordered set which groups and classifies education programmes according to the knowledge, skills, competencies and qualifications which they are designed to impart.

#### Economic activity (ACTIVITY)

Based on the ISIC and NACE classifications. Combination of actions that result in the production, distribution and consumption of goods or services.

#### Occupation (OCCUPATION)

Based on the ISCO classification. Job or position held by an individual who performs a set of tasks and duties.

#### Degree of urbanisation (DEG\_URB)

Degree to which area of residence is urbanised.

#### Seasonal Adjustment (SEASONAL\_ADJUST)

Statistical technique used to remove the effects of seasonal and calendar influences operating on a data series.

#### Migrant status (MIGRANT\_STATUS)

Distinction between migrants and non-migrants.

#### Origin of migrant (MIGRANT\_ORIGIN)

Refers to the country of origin of a migrant.

#### Nationality (NATIONALITY)

Refers to the country of nationality of a person.

#### Type of Production Unit (PROD\_UNIT)

Refers to the nature of the economic unit in which a job is located as determined by its legal organization, type of accounts keep and orientation of the output. For example, Informal, Formal, Household.

#### Number of persons engaged in the economic unit (SIZECLASS)

Refers to the number of employed persons engaged in the economic unit in which a job is located.

#### Earnings case (EARNINGS\_CASE)

Household composition and income situation of average workers at national level.

#### Time period (TIME\_PERIOD)

Time span or point in time to which the observation actually refers. The observation corresponds to a specific point in time (e.g. a single day) or a period (e.g. a month, a fiscal year, or a calendar year). This is used as a time stamp and is of particular importance for time series data.

#### Frequency (FREQ)

Time interval at which observations occur over a given time period. If a data series has a constant time interval between its observations, this interval determines the frequency of the series (e.g. monthly, quarterly, yearly).

#### Composite breakdown (COMPOSITE\_BREAKDOWN)

A single dimension that is used to hold multiple breakdowns which are seldom used and not in the same data exchange or series. Its main purpose and advantage is that when a DSD has so many concepts it may be difficult to implement, this dimension replaces many concepts with only two concepts: this dimension and the Composite breakdown in use (COMPOSITE\_BREAKDOWN\_PFX) attribute.

This dimension must be used in conjunction with the “Composite breakdown in use” attribute that indicates the breakdown that is being used. For example, if the data is broken down by “Type of employment sought: Only full-time paid job” then the data coding includes:

* **CL\_COMPOSITE\_BREAKDOWN\_PFX**: EMPSGT (meaning the breakdown is “Type of employment sought”)
* **CL\_COMPOSITE\_BREAKDOWN**: EMPSGT\_FT (meaning the breakdown value is “Only full-time paid job”)

See the [composite breakdown series key coding example](#_Example_of_coding) for its use in a dataflow.

## Attributes

See the “Concept Scheme” worksheet in the “Labour Stats DSD Matrix” workbook for exact details on attachment level, optionality, data type, etc.

#### Currency (CURRENCY)

Monetary denomination of the object being measured.

#### Confidentiality status (CONF\_STATUS)

Information about the confidentiality status of the object to which this attribute is attached.

#### Decimals (DECIMALS)

Number of digits of an observation to the right of a decimal point.

#### Observation status (OBS\_STATUS)

Information on the quality of a value or an unusual or missing value.

#### Composite breakdown in use (COMPOSITE\_BREAKDOWN\_PFX)

Used in conjunction with the composite breakdown dimension to provide the prefix corresponding to the breakdown used. See the “[Composite breakdown](#_Composite_breakdown)” entry for more details.

#### Time format (TIME\_FORMAT)

Technical format for the representation of the time concepts in the SDMX message. Both it and its related codelist are part of the technical standards for SDMX-EDI and SDMX-XML.

#### Unit multiplier (UNIT\_MULT)

Exponent in base 10 specified so that multiplying the observation numeric values by 10^UNIT\_MULT gives a value expressed in the unit of measure.

#### Data compilation (DATA\_COMP)

Operations performed on data to derive new information according to a given set of rules.

#### Detailed description of the group of series (COMMENT\_TS)

Descriptive text which can be attached to the time series.

#### Comments to the observation value (COMMENT\_OBS)

Descriptive text which can be attached to the observation.

#### Comments on dataset level (COMMENT\_DSET)

Descriptive text which can be attached to the dataset.

#### Time period collection (TIME\_PER\_COLLECT)

#### Segment(s) of the time period for which the observations have been collected (such as middle, average or end of period) for the target reference period. This is not a stand-alone concept; it is related to the concept TIME\_PERIOD which it further specifies (e.g. TIME\_PERIOD "Monthly" and TIME\_PER\_COLLECT "End of period").

#### Organisation disseminating the data being reported (DISS\_ORG)

This metadata element is needed in order to differentiate the compiling organisation from the organisation disseminating the data. The dissemination agency could be different from the reporting agency and the compilation agency.

#### Compiling organisation (COMPILING\_ORG)

The compiling organisation.

#### Embargo date and time (EMBARGO\_TIME)

#### Exact time at which the data can be made available to the public. Usually, there is a time delay between the finalisation of the production process of statistical data and the moment when the data produced are released and made available to the users. This point in time where data are made publicly available is called "embargo time".

#### Base period (BASE\_PER)

Period of time used as the base of an index number, or to which a constant series refers.

#### Pre-break value (PRE\_BREAK\_VALUE)

Observation, at a time series break period, that was calculated using the old methodology.

#### Reason for break in series (BREAK\_REASON)

To identify reason for break in series.

#### Maximum age (COVERAGE\_MAXAGE)

Maximum age of the population of this series.

#### Minimum age (COVERAGE\_MINAGE)

Minimum age of the population of this series.

#### Ownership sector (COVERAGE\_OWN)

The ownership sector.

#### Geographical coverage (COVERAGE\_GEO)

Characterisation of the statistical units according to geographical criteria.

#### Economic or other sectors covered by the statistics (COVERAGE\_ACTIVITY)

This metadata element lists the sectors covered by the data set produced These sectors can be institutional sectors, economic or other sectors (e.g. local government, agriculture, forestry, or business services).

#### Economic unit size coverage (COVERAGE\_SIZECLASS)

Economic unit size coverage.

#### Measure notes 1 to 5 (MEASURE\_NOTES1 to MEASURE\_NOTES5)

Multiple concepts that allow the provision of measure-related coded metadata.

#### Title of the statistical source (TITLE\_STAT\_SOURCE)

Title of the statistical source.

#### Title (TITLE\_TS)

Textual label used to refer to a statistical object on the time series.

# The DSD Matrix workbook

## Overview

The Matrix file summarises all concepts and code lists contained in the reporting framework. Those have been compiled from the Labour statistics transmission programmes. It contains the following worksheets:

* **Indicator-Concept matrix sheet**: describes how the information in the existing Labour indicators has been “decomposed” into separate concepts;
* **Concept Scheme sheet**: describes the concept scheme and details of the concepts;
* **DSDs**: Describes the DSDs in the “Global DSD for Labour” framework, their version and agency, etc.
* **DSD-Concept Matrix sheet**: describes the relationship between the Labour data flows, the labour statistics concepts, and the DSD/data flow associations;
* **Codelist sheets**: describes the contents of each code list. (pending) The codelists sheets also define the constraints for each dataflow/concept scheme.

## Indicator-Concept Matrix Worksheet

This worksheet contains the mapping between the Labour indicators that are known in several existing exchange frameworks, and the concepts in the Labour Global DSD concept scheme. The advantages of decomposing the labour indicators (as opposed to staying with the current list) are:

* Define a single exchange framework between agencies for labour statistics;
* Standardise and reduce the labour statistics variables between organisations, rather than having many variations on the same variable for specific organisations;
* Make it easier to exchange new statistics; for example, a combination of existing codes can be used rather than creating a new indicator;
* Make it easier to map from the Labour Global DSD to internal databases: The disambiguation of variables should make it easier to map to many internal databases in different agencies;
* Avoid frequent maintenance changes to the Labour reporting framework (global DSD): A new indicator would mean a new versioning of the DSDs would have to be created and disseminated.

This table shows a part of the matrix



The legacy labour indicators are in the left green-shaded column. The information in the indicators has been “decomposed” into the right-hand concepts (in white) that are part of the Labour concept scheme.

The cells in each concept column (the matrix) show how the concept is coded for the indicator. For example, the original indicator **“Median monthly nominal earnings**” is now coded as:

* **UNIT\_MEASURE**: Currency per person
* **MEASURE**: Earnings
* **AGGREGATION\_OPERATION**: Median
* **INDICATOR\_REF\_PER**: Monthly
* **LABOUR\_FORCE\_STATUS**: EMP

A cell with a **%** sign means that several codes may be used for the measure – for example, LABOUR\_FORCE\_STATUS has a % for the last measure “Labour cost”, so several values in the code list can be used, but not all (the allowed set of values is defined in a constraint).

## DSD-Concept Matrix Sheet

### Dataflow/Concept coding

The list of concepts in the concept scheme is shown in the columns starting with FREQ. Each coded concept has a hyperlink allowing easy navigation to its code list sheet. The matrix also links a data flow (row) to a DSD.



The cells under each concept summarise how the concept is used and constrained for the data flow. The actual constraint is defined in each Code List. The matrix cells contain:

* A **blank** if the concept is not used in the data flow;
* A **#** sign if the concept is used and any code is valid (there is no constraint);
* A **code** if the concept is fixed to a single code (constrained);
* A **%** sign if the concept is used but a subset of codes are valid (constrained).

For example, the dataflow “Labour Cost Index – Weight” uses the DSD:LCI (Labour Cost Index) and is constrained in this way:

* **FREQ**:A (Annual)
* **REF\_AREA**: # (any code is valid)
* **MEASURE**: % (constrained to a subset of codes)
* **UNIT\_MEASURE**: % (constrained to a subset of codes)
* **ACTIVITY**: % (constrained to a subset of codes)
* **SEASONAL\_ADJUST**: \_Z (constrained to this code)

### Dataflow Sender/Receiver

For each dataflow, the sender and receiver agency of the data. This is used partly to decide the grouping of the dataflows into DSDs, and for information:



## Code Lists

Each code list is documented on a separate worksheet. Generally, the Code Value and Description is shown. Some code lists have extra elements such as Integrity Rules that describe how the code should be validated. This is an example of some codes in CL\_LABOUR\_FORCE\_STATUS:

|  |  |  |  |
| --- | --- | --- | --- |
| **Code Value** | **Code Description** | **Parent Code** | **Integrity rule** |
| WAP | Working age population |  | LF+OLF |
| ELF | Extended labour force |  | LF+PLF |
| PLF | Potential labour force |  | SNA+NSA |
| LF | Labour force |  |  |
| EMP | Employment | LF |  |
| EMP\_TRU | Time-related underemployment | EMP |  |
| UNE | Unemployment | LF |  |
| NE | Not in employment |  | OLF+UNE |

The integrity rule for the code ELF shows that it should be the sum of codes LF and PLF. This code list has a hierarchy that is shown by the **Parent Code** column; for example, **Employment** is a child of **Labour force**.

# Example coding

The following are some examples of data series keys for the dataflows in the Labour Global DSDs. Please note that the series key combinations are purely synthetic and should not be taken as de-facto implementations.

**DSD**: Job Vacancy Statistics **Dataflow**: Quarterly Vacancy Data with Economic Activity Breakdown

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **FREQ** | **REF\_AREA** | **MEASURE** | **UNIT\_MEASURE** | **ACTIVITY** | **SEASONAL\_ADJUST** | **OCCUPATION** | **SIZECLASS** |
| Q | AT221 | JVR | IDX | ISIC2\_11 | S | ISCO88\_3212 | P10T19 |

**DSD**: Labour Cost Index **Dataflow**: Labour Cost Index – Weight

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FREQ** | **REF\_AREA** | **MEASURE** | **UNIT\_MEASURE** | **ACTIVITY** | **SEASONAL\_ADJUST** |
| A | FR | LCI\_WAA\_LC | CUR\_PER | ISIC2\_2100 | \_Z |

**DSD**: Labour Force **Dataflow**: Employees

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FREQ** | **REF\_AREA** | **MEASURE** | **UNIT\_MEASURE** | **LABOUR\_FORCE\_STATUS** | **ACTIVITY** | **SEX** |
| **SEASONAL\_ADJUST** | **WORKER\_STATUS** | **INDICATOR\_REF\_PER** | **AGE** | **AGGREGATION\_OPERATION** | **CURRENCY** | **HOURS\_TYPE** |
| **OCCUPATION** | **VALUE\_TYPE** | **EDUCATION\_LEV** | **DEG\_URB** | **SIZECLASS** | **MIGRANT\_ORIGIN** | **NATIONALITY** |
| **SOURCE\_TYPE** | **PROD\_UNIT** | **MIGRANT\_STATUS** | **COMPOSITE\_BREAKDOWN** |  |  |  |
| A | AU | EES | PS | EMP | ISIC2\_130 | \_T |
| N | ICSE93\_1\_PERM | \_Z | Y25T29 | \_Z | \_Z | USUAL |
| ISCO88\_6 | \_Z | \_Z | \_Z | \_Z | \_Z | \_Z |
| HS\_LFS | \_Z | \_Z | \_Z |  |  |  |

## Example of coding with Composite Breakdown

This series shows the use of the Composite Breakdown dimension (COMPOSITE\_BREAKDOWN) to describe the “type of incapacity” from the injury. The dimension requires the use of the attribute Composite breakdown in use (COMPOSITE\_BREAKDOWN\_PFX) which is highlighted in red below.

**DSD**: Labour Force **Dataflow**: Occupational injuries

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FREQ** | **REF\_AREA** | **MEASURE** | **UNIT\_MEASURE** | **LABOUR\_FORCE\_STATUS** | **ACTIVITY** | **SEX** |
| **SEASONAL\_ADJUST** | **WORKER\_STATUS** | **INDICATOR\_REF\_PER** | **AGE** | **AGGREGATION\_OPERATION** | **CURRENCY** | **HOURS\_TYPE** |
| **OCCUPATION** | **VALUE\_TYPE** | **EDUCATION\_LEV** | **DEG\_URB** | **SIZECLASS** | **MIGRANT\_ORIGIN** | **NATIONALITY** |
| **SOURCE\_TYPE** | **PROD\_UNIT** | **MIGRANT\_STATUS** | **COMPOSITE\_BREAKDOWN** | **COMPOSITE\_BREAKDOWN\_PFX** |  |  |
| A | AU | INJ\_F | CASE | \_Z | ISIC2\_1110 | F |
| N | ICSE93\_1\_PERM | \_Z | Y25T29 | \_Z | \_Z | USUAL |
| ISCO88\_6 | \_Z | \_Z | \_Z | \_Z | \_Z | \_Z |
| HS\_LFS | \_Z | NOMIGRANT | INCTYP\_PERM | INCTYP |  |  |

# Availability of the SDMX Artefacts

## SDMX Artefacts

The SDMX artefacts will be generated during the public review and made available on the SDMX.org public review site. Once in production, the [SDMX Global Registry](https://registry.sdmx.org/FusionRegistry/overview.html) will be used as a central reference for the concept scheme, Data Structure Definitions, and related objects.

# Governance

## Ownership Group

A group formed of representatives from the International Labour Organisation, OECD, Eurostat, ECB, and the World Bank are responsible for approving changes to the Global DSD, its documentation and supporting material, the procedures outlined in this document, and are designated the “Global DSD for Labour Statistics ownership group” (OG).

The OG reacts to proposals from their internal organisations, respective constituencies, and related statistical communities.

## Maintenance Agencies

The SDMX artefacts in and referenced by the “Global DSD for Labour Statistics” model have separate maintenance agencies (MAs). An MA is not a decision-making body; instead, an MA reacts to decisions taken by the OG (see [above](#_Ownership_Group)) that affect the artefacts it is responsible for, and executes any approved changes following the agreed procedures according to the maintenance schedule.

Here are the MAs for the artefacts in the Global DSD:

* **Concept Scheme and concepts**: ILO
* **DSDs**: ILO
* **Data flows and Constraints**: the receiving organisation of the Data flow in the DSD-Concept Matrix
* **Codelists**: as stated in the Concept Scheme worksheet

If the Labour DSDs require a Code List that has already been defined by another project, it is considered to reference the existing Code List rather than recreate it, especially the “cross-domain Code Lists”. The MA of several Code Lists reflects this, for example the CL\_SEX code list is maintained by “SDMX”, meaning that it is a cross-domain code list maintained by the SDMX Statistical Working Group.

# Maintainence Schedule

To be defined