

Pilot Guidelines for Global DSDs on Price Statistics: CPI/HICP

- 1 Introduction, governance and scope
- 2 Content review information
 - 2.1 Implementation Model
 - 2.2 General remarks
 - 2.3 Dimensions and attributes
 - 2.4 DSD matrix
- 3 Technical review information
 - 3.1 Link between the DSD matrix and the SDMX artefacts
 - 3.2 SDMX artefacts
- 4 Coding examples
- 5 Annex
 - 5.1 Annex 1 - Data Flow documentation

Introduction, governance and scope

The Ownership Group for SDMX implementation for macro-economic statistics (SDMX-MES OG) mandated a technical sub-group to propose global Data Structure Definitions (DSDs) for the implementation of SDMX in Price Statistics. The SDMX-MES OG is already responsible for SDMX DSDs in the areas of National Accounts (incl. Government Finance), Balance of Payments and Foreign Direct Investment. More information on the mandate of the SDMX-MES OG and the existing packages can be found on the SDMX website:

SDMX-MES package on National Accounts and Government Finance: https://sdmx.org/?page_id=1498

SDMX-MES packages on Balance of Payments and Foreign Direct Investment: https://sdmx.org/?page_id=1747

The sub-group on SDMX for price statistics is composed of representatives from BIS, ECB, Eurostat, ILO, IMF, OECD and UNSD. The overall scope of the work has been defined to cover the area of price indexes, covering the following 5 sub-domains:

1. Consumer Price Index (national CPI and HICP)
2. Producer Price Index
3. Import and Export Price Index
4. Residential Property Price Index / Indicators
5. Commercial Property Price Index / Indicators

Price levels or other price related statistics as well as PPPs are currently not in scope. Work in the sub-group is conducted in several phases. It was agreed to start with the area of CPI/HICP and extent to other sub-domains at a later stage. The DSDs and related materials should cover the whole statistical process as far as price indexes are concerned: from reporting and collection from national to international organisations to public dissemination and data sharing. This includes some reference metadata which is "close to the data" and is usually transmitted together with data. Other more high level reference metadata (e.g. methodological notes) are currently out of scope. This aspect would be picked up once discussion on Metadata Structure Definitions (MDSs) is elaborated for all domains within the MES area.

The group has defined a "DSD Matrix" as an inventory for structuring the relevant data flows in the area of price statistics. The DSD Matrix includes the list of concepts that are necessary to code the reporting requirements of the different SDMX sponsors agencies. The codification of domain concepts in the DSDs has to be descriptive and comprehensive enough to fulfil any current and future data exchange requirements within the scope of price indexes. In this note, a short description will be provided of the contents of a DSD. Also the before mentioned DSD Matrix will be explained in some more detail. Doing so, we will also try to describe the main terms and definitions used.

The process follows the generic "Checklist for SDMX Design Projects" that is available on the SDMX website: https://sdmx.org/?sdmx_news=checklist-for-sdmx-design-projects. An important milestone in every SDMX project is the pilot review phase. It consists of two phases: a content review and a technical review. The aim of the content review is that the project's *DESIGN* stage material is understandable, well-designed, and that it covers the scope of the project. The main deliverable for review is the "DSD matrix". The aim of the technical review is to test that the project's SDMX artefacts can be implemented at system level as efficiently as possible. It includes to test material such as example data messages and xml-schemas. The SDMX artefacts should be available in a registry, and the implementation & usage guidelines are sent for documentation. Further details on the process can be found in the checklist mentioned above.

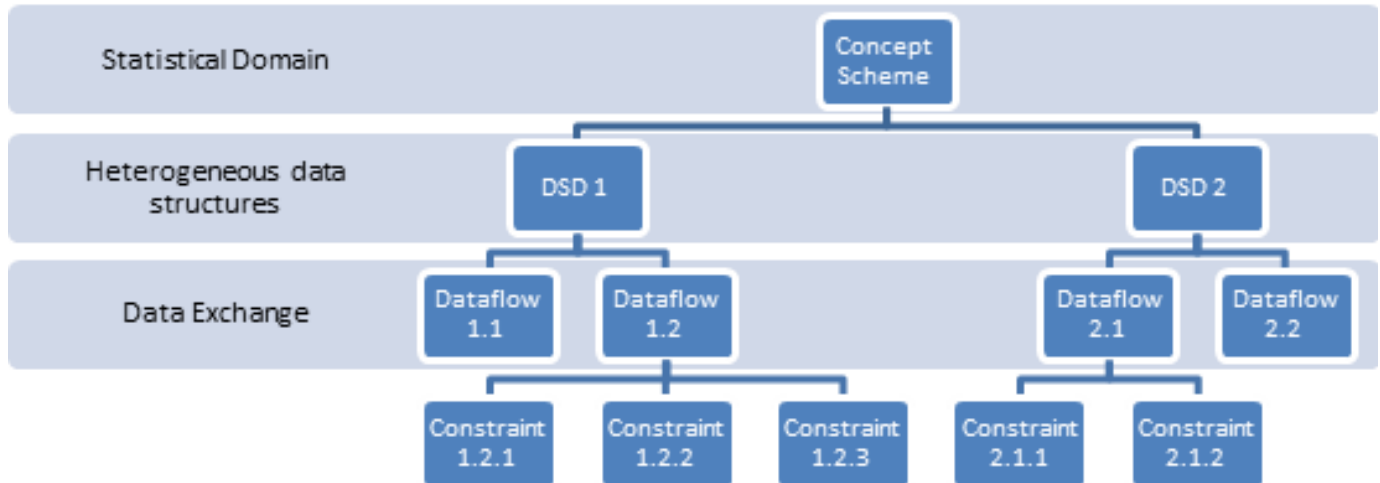
For this project it was decided to conduct the content and technical reviews in parallel. This guideline document should support participants conducting the pilot review. Please send any comments you may have on the package to ESTAT-HICP@ec.europa.eu and Daniel.Suranyi@ec.europa.eu by **31st March 2018**. The group will review the comments received during the pilot consultation and publish decisions on any follow-up actions taken and created an updated version of the package as needed in due time after that.

Content review information

Implementation Model

The model chosen to implement SDMX for price indexes follows the basic model used also in SDMX for National Accounts. The approach is also

expressed in the guidelines for "Modelling a Statistical Domain for Data Exchange in SDMX" available from the SDMX Statistical Working Group: https://sdmx.org/?page_id=4345#Modelling. The document suggests a generic implementation model pictured below:



Following this generic model, the outcome of the design project so far are:

- 1 Concept Scheme covering the area of price indexes for CPI and HICP data flows
- 2 Data Structure Definitions: CPI/HICP (consumer price index and EU harmonised CPI) and HICP AP (HICP administered prices)

Data Flow and Constraint artefacts are not yet created for the pilot review. Further mapping and creation of artefacts will only be done once the overall model is approved. Some examples are provided as part of the pilot package.

General remarks

Note that as done in the other DSDs within macro-economic statistics, we work with a single measure (OBS_VALUE). No multiple measures are used. Thus only dimensions and attributes are described below. For the coded dimensions and attributes, code lists have been taken as far as available in that order from:

- existing cross-domain code lists
- existing code lists from other MES DSDs
- derived from existing classifications
- created by the group

For the last two cases, the guidelines for the Creation and Management of SDMX Code Lists have been used (https://sdmx.org/?page_id=4345#CodeListGuideline). Some existing code lists have been extended. In order not to impact the existing DSDs, they have been created in the pilot registry with ESTAT maintenance agency and a different version number for the purpose of the pilot. Once approved, the original Code Lists would be updated and used across all MES DSDs. These are marked in red in the DSD matrix for "draft maintenance agency / version for pilot use".

Dimensions and attributes

Within a data message, dimensions are used to uniquely identify the observation. They are always mandatory. Attributes can be used to further describe the data. They can be linked to several levels (dataset, series, observation) and can be mandatory or optional. All concepts can be coded (linked to a code list) or be defined through a data type (e.g. String, DateTime, ...).

DSD matrix

The Matrix file summarised all concepts and code lists contained in the reporting framework. Those have been selected from various sources along the data flows documented in annex 1.

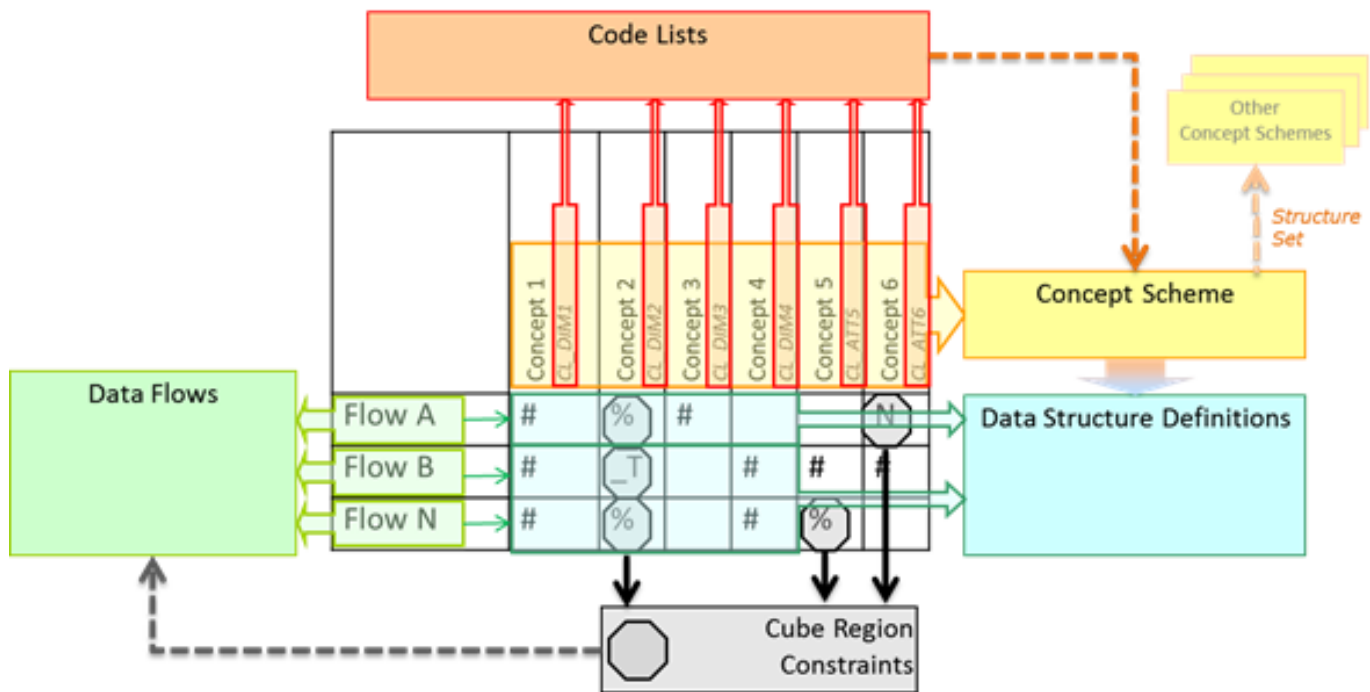
It contains the following sheets:

- **Overview DSDs**, showing the concepts used, their representation and the link to the DSDs;
- **Matrix sheet**, kept as a placeholder for showing the relationship between the data flows and concepts once implemented;
- **Code list sheets**, showing the contents of each of the code lists used and some additional comments like for instance integrity rules. Code lists marked in green are sourced from cross-domain code lists; code lists marked in yellow are MES shared code lists. Note that for some shared code lists, several codes on top are marked in yellow as well. Those are suggested additions which will be merged into the existing code lists upon completion of the pilot.

Technical review information

Link between the DSD matrix and the SDMX artefacts

Following the generic implementation model, the DSD matrix is linked to the SDMX artefact as follows:



For the pilot review, the matrix documents the Concepts Scheme, the DSDs and the Code Lists. Data Flows have been used to analyse the Concepts and Code Lists but will only be fully documented after the pilot review (see also annex 1 for data flows included for final implementation).

SDMX artefacts

In most cases, dimensions in SDMX DSDs are coded. In fact for all MES DSDs existing so far, all dimensions have been coded. For modelling CPI/HICP, it was decided to define the BASE_PER dimension with the data type *ObservationalTimePeriod* in order to be able to express flexibly any period as base period for the index. This should get special attention during the pilot review, to prepare systems for mapping a dimension to a data type, in case this was not yet tested in the specific environment used by the participant.

The pilot artefacts can be access through the Fusion Cloud registry (<https://registry.sdmxcloud.org/>). Once approved, the final artefacts will be published in the Global SDMX Registry (<https://registry.sdmx.org/>). Note that the interfaces look different, but both registries allow user interface as well as web service access. The DSD identifiers and URLs to the pilot DSDs are the following:

Pilot maintenance agency	DSD ID	Pilot DSD version	REST URL (SDMX 2.1, all references)
ESTAT	CPI	0.2	https://registry.sdmxcloud.org/ws/public/sdmxapi/rest/datastructure/ESTAT/CPI/0.2/?format=sdmx-2.1&detail=full&referen
ESTAT	HICPAP	0.2	https://registry.sdmxcloud.org/ws/public/sdmxapi/rest/datastructure/ESTAT/HICPAP/0.2/?format=sdmx-2.1&detail=full&rel

Note that the maintenance agency for the DSDs is currently Eurostat for purposes of the pilot only. The final maintenance agency for the DSDs might change depending on Ownership Group agreement after the pilot and before the DSDs are finally published in the global DSD registry. Other parameters of the REST URL can be modified as needed by the implementer.

Coding examples

Annex

Annex 1 - Data Flow documentation

The following data flows have been used to analyse the requirements for SDMX Concepts:

