*Data provisioning*

A provisional agreement is based on Data Structure Definitions (DSDs), defined by data collector in agreement with data provider, representing specific information to data provider on which data must be exchanged.

A DSD, can be further restricted using constraints in dataflows, for example data in dataflows can be organised by frequency so that, also referring to the same DSD in which the frequency assumes different values (e.g. M, A, W), dataflows can be distinguished assigning a specific value to the frequency concept so to exchange monthly, annual or weekly data using different dataflows. The possibility to “restrict” the whole domain of a DSD is implemented in the standard, using *constraints*. Through constraints is possible that a set of concepts assumes a specific combination of values (*key set* constraint) or that one or more concepts take values from a sub set of the related code lists (*cube region* constraint).

In the Matrix of chapter 2.4 of the Guideline: “MODELLING STATISTICAL DOMAINS IN SDMX”, constraints are represented in a Matrix.

Below an example for a key set constraint in which, for the dataflow Flow N, the code A1 of CL\_DIM2 for Concept 2, can be combined only with the codes {B1, B2} of code list CL\_DIM5 for Concept 5 while the code A2 can be associated only with codes B3 and B5 and A3 and A4 only with B4. The blank code instead, must be replaced with a code representing the no usage of the concept (e.g. NA= not applicable).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Concepts →  ↓ Dataflows | Concept 1 *CL\_DIM1* | Concept 2 *CL\_DIM2* | Concept 3  *CL\_DIM3* | Concept 4  *CL\_DIM4* | Concept 5  *CL\_DIM5* | Concept 6  *CL\_DIM6* |
| Flow A | # | {A1, A2, A3 } | # | NA | NA | N |
| Flow B | # | \_T | NA | # | # | # |
| … |  |  |  |  |  |  |
| Flow N | # | A1 | NA | # | {B1, B2} | # |
| # | A2 | NA | # | {B3, B5} | # |
| # | {A3, A4} | NA | # | {B4} | # |
| *Legend* | *# concept fully used in dataflow*  *(code) concept fixed to single code in dataflow* | | | | | |

***Key set in a matrix***

In the example for the cube region constraint, the concept2 of Flow N can only assume values {A1, A2, A3, A4} of the complete codelist CL\_DIM2 while the concept5 can assume only values {B1, B2, B3, B4, B5} of the codelist CL\_DIM5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Concepts →  ↓ Dataflows | Concept 1 *CL\_DIM1* | Concept 2 *CL\_DIM2* | Concept 3  *CL\_DIM3* | Concept 4  *CL\_DIM4* | Concept 5  *CL\_DIM5* | Concept 6  *CL\_DIM6* |
| Flow A | # | {A1, A2, A3 } | # | NA | NA | N |
| Flow B | # | \_T |  | # | # | # |
| … |  |  | NA |  |  |  |
| Flow N | # | {A1, A2, A3, A4} | NA | # | {B1, B2, B3, B4, B5} | # |
| *Legend* | *# concept fully used in dataflow (code) concept fixed to single code in dataflow* | | | | | |

***Cube region in a matrix***

Constraints in version 2.0 of SDMX are Maintainable artefacts so that they are identified through an id, agency and version. They can be stored independently for the artefacts on which they are applied allowing a better reusability.

There is another way to restrict data in a dataflow through the use of the partial code list that is not explicit declared like in the previous examples. It consists in sending to different data providers different subset on the same codelists by denoting them as “partial”.

This constraint has no impact on the definition of DSD neither on the dataflow one, it only represents an indication for a data provider receiving it, to disseminate only codes received in the partial code list. This has been applied in the Census hub project to the geographical code list sent to data providers partially and including only the specific codes for the related data provider territory.