

- หลักเกณฑ์พื้นฐานของการสร้างรายการรหัส (Basic principles for creation of SDMX code list)
- คุณลักษณะพื้นฐานของรายการรหัส (Basic characteristics of SDMX code list)
- การกำหนดพื้นฐานของรายการรหัส (Basic criteria of SDMX code list)

## หลักเกณฑ์พื้นฐานของการสร้างรายการรหัส (Basic principles for creation of SDMX code list)

Code list	
code list name	Sex
code list id	CL_SEX
code list version	1.0
code list agency	SDMX

นิยาม : รายการที่ถูกนิยามไว้จากแนวคิดทางสถิติ

code list ประกอบด้วย 3 ส่วนหลัก คือ

1. id
2. version number
3. reference to maintenance agency

Code	
code id	F
code name	Female
code id	M
code name	Male

code ประกอบด้วย 2 ส่วนหลัก คือ

1. id
2. name

หมายเหตุ code list name คือคำอธิบายหัวข้อของ code list id

## หลักเกณฑ์พื้นฐานของการสร้างรายการรหัส (Basic principles for creation of SDMX code list)

### □ code list ประกอบด้วย 3 ส่วนหลักคือ

1. Code list id : กำหนดให้ขึ้นต้นด้วย CL\_ เพื่อระบุความชัดเจน  
ใช้ตัวอักษร A – Z หรือตัวเลข 0 – 9 ต้องเป็นตัวใหญ่  
\_ ขีดล่าง (Underscore) ใช้สำหรับการเชื่อม code

ตัวอย่างเช่น Code list id : CL\_SEX  
CL\_EDUCATION\_LEVEL  
CL\_AREA

## หลักเกณฑ์พื้นฐานของการสร้างรายการรหัส (Basic principles for creation of SDMX code list)

2. Code list version : มีข้อกำหนดในการตั้งตาม SDMX official website : [https://sdmx.org/?page\\_id=4345#Versioning](https://sdmx.org/?page_id=4345#Versioning)

ตัวอย่างเช่น เมื่อมีการกำหนด code list ครั้งแรกของหน่วยงานควรจะใช้ version 1.0

Syntax : MAJOR . MINOR . PATCH

ตัวอย่างที่ 1 – MAJOR เปลี่ยน :

Id	Old Name	New Name	Old version	New version
CP01115	Other products	Pizza and quiche	1.0	2.0

ตัวอย่างที่ 2 – MINOR เปลี่ยน :

Id	Change type	Old version	New version
CL_OBS_STATUS : Code List	Addition of a new code X	1.0	1.1

ตัวอย่างที่ 3 – PATCH เปลี่ยน : PATCH แทบไม่ค่อยพบการเปลี่ยนใน SDMX

Id	Old Name	New Name	Old version	New version
PRODUCT_TO	Product classification	Product classification (input-output product*product)	1.2	1.2.1

## หลักเกณฑ์พื้นฐานของการสร้างรายการรหัส (Basic principles for creation of SDMX code list)

3. Code list agency : เป็นหน่วยงานที่กำหนด code list

ตัวอย่างเช่น Code list agency : LIO (International Labour Organization)

ESTAT (Statistical Office of the European Union)

IMF (International Monetary Fund)

❑ code ประกอบด้วย 2 ส่วนหลักคือ

1. Code id : สามารถกำหนดเป็นตัวอักษร (A-Z) ตัวเลข (0-9) หรือสัญลักษณ์ก็ได้

ตัวอย่างเช่น Code id : F M \_T M2\_3

2. Code name : กำหนดให้ใช้ตัวอักษรไม่เกิน 254 ตัวอักษร

ตัวอย่างเช่น Code name : Female Male Total Two and three months

## คุณลักษณะพื้นฐานของรายการรหัส (Basic characteristics of SDMX code list)

- code list สามารถอ้างอิงได้หลากหลายแนวคิดทางสถิติแต่ละพื้นที่ และควรอธิบายรายละเอียดให้ถูกต้อง ครบถ้วน
- code ควรนิยามที่ชัดเจน ซึ่ง code ที่แตกต่างกันไม่ควรมีความหมายเดียวกัน  
ตัวอย่างเช่น ยุโรปที่รวมกรีนแลนด์ code id คือ EG ส่วนยุโรปที่ไม่รวมกรีนแลนด์ code id คือ EG\_OTHER
- ควรหลีกเลี่ยงการใช้ code id ที่ซ้ำกันถึงแม้ว่าจะเป็นคนละ version เนื่องจากระบบฐานข้อมูลมีข้อมูลเดิมอยู่  
ตัวอย่างเช่น code id คือ CS - Czechoslovakia was suppressed in 1993  
CS - Serbia and Montenegro in 2003

## การกำหนดพื้นฐานของรายการรหัส (Basic criteria of SDMX code list)

- code list ให้มีความสอดคล้องตามหลักมาตรฐานสากลที่มีการตกลงกันไว้ ISO UN และองค์กรประเทศต่าง ๆ
- การกำหนด code โดยทั่วไป

ตัวอย่างเช่น \_S หมายถึง บางส่วน/ส่วนย่อย (Subtotal)

\_T หมายถึง ทั้งหมด (Total)

\_U หมายถึง ไม่มีข้อมูล (No data/unknown)

\_N หมายถึง ไม่ได้รับการตอบสนอง (Non response)

- การใช้กำหนด code ที่ต่อท้าย (suffixes) เพื่อเพิ่มจำนวนลำดับชั้น แต่ควรระวังในการใช้

ตัวอย่างเช่น - M\_S อาจจะแปลความหมายได้ 2 แบบคือ หมวด M และ S หรือ หมวด M ที่รวมบางส่วน

แต่ควรจะสร้างเป็น M\_SUB หรือ M\_SUBTOTAL

- PLANTS\_S1, PLANTS\_S2, CATTLE\_S1

- PLANTS\_O, CATTLE\_O

## ตัวอย่าง การสร้างรายการรหัส (code list) และรหัส (code)

Code list	
code list name	Age classes
code list id	CL_AGE_CLASSES
code list version	1.0
code list agency	NSO

Code id	Code name
D	day
W	week
M	month
Y	year
T	expressing ranges
_	combination of two codes
X	"except" or "excluding"
GT	greater than
LT	less than
GE	equal to or greater than
LE	equal to or less than

Code id	Code name
Y15T20X16	from 15 years to 20 years excluding 16 years
Y_GT30	over 30 years
Y_LT50	less than 50 years
M2_3	two and three months
D_GE4	four days or over
Y_LE3	three years or less



## GUIDELINES FOR THE CREATION AND MANAGEMENT OF SDMX CODE LISTS

(Version 3.0 19/01/2018)

**แนวทางการสร้างรายการรหัส (code list) SDMX** เพื่อนำไปใช้ในกระบวนการทางสถิติของธุรกิจ โดยเฉพาะเมื่อ SDMX อยู่ในโดเมน (domain) สถิติ โดยรายการรหัสจะถูกสร้างอยู่ในกลุ่มที่มีความหมายที่เกี่ยวข้องเป็นระบบ และมีมาตรฐานเดียวกัน โดยรายการรหัสจะนำไปใช้ในการรวบรวม แลกเปลี่ยน และจัดการข้อมูล ซึ่งการจัดข้อมูลมีทั้งการรวบรวมให้อยู่ในกลุ่มเดียวกัน และการแยกกลุ่มเพื่อให้มีความหมายสามารถนำไปใช้ในการวิเคราะห์ที่มีความซับซ้อน นอกจากนี้แล้วยังสามารถนำไปใช้ในการนำเสนอข้อมูลให้มีมาตรฐาน เพื่อนำไปใช้ในการออกนโยบายหรือการตัดสินใจ และเพื่อให้กระบวนการวัดผลมีมาตรฐาน

**หลักเกณฑ์เบื้องต้นรายการรหัส (code list) และรหัส (code)** คือรายการรหัส (code list) จะประกอบด้วยส่วนที่บังคับ (mandatory) สามส่วน คือ id version number และ agency ส่วนรหัส (code) จะประกอบด้วย สองส่วนหลัก คือ id และ name ซึ่งคำบรรยาย (description) นั้นถือว่าเป็นทางเลือก (option) ในการอธิบายรายละเอียดเพิ่มเติม ส่วนนี้สามารถนำเสนอได้ในภาษาที่หลากหลาย โดยเบื้องต้นจะใช้ภาษาอังกฤษเป็นหลัก

**ลักษณะพื้นฐานของรายการรหัส (code list) และรหัส (code) เช่น**

- รายการรหัสสามารถอ้างอิงได้จากแนวคิดทางสถิติที่หลากหลาย
- รหัสควรมีการนิยามที่ความชัดเจน เช่น ยุโรปที่รวมกรีนแลนด์ (Greenland) จะต้องมีการ id ที่แตกต่างกันกับยุโรปที่ไม่รวมกรีนแลนด์ และรหัสที่แตกต่างกันไม่ควรจะหมายถึงสิ่งเดียวกัน
- รหัสควรมีความยืดหยุ่นในการที่จะเพิ่มลำดับชั้น (hierarchies)

โดยลักษณะของ code อาจจะต่างกันเนื่องจากมีข้อจำกัดที่ขึ้นอยู่กับผู้ใช้งานทั้งสองฝ่ายที่จะระบุ หรืออธิบายรายละเอียดของข้อมูลให้มีความครอบคลุมอย่างชัดเจนของข้อมูลนั้น ๆ เพื่อการแลกเปลี่ยนข้อมูลที่ต้องการตามมาตรฐานเพื่อให้ผู้ใช้งานมีความเข้าใจตรงกัน

**เกณฑ์พื้นฐานของการพัฒนารายการรหัส (code list)** สิ่งสำคัญคือ รหัสควรมีความคงตัวต่อมาตรฐานสากล ต้องเป็นไปตามมาตรฐานเดิมที่ตกลงกันไว้ ISO UN และองค์กรประเทศต่าง ๆ และการกำหนดตามมาตรฐานข้อตกลงเพิ่มเติมโดยขึ้นอยู่กับข้อจำกัดและความต้องการของผู้ใช้งานข้อมูลนั้น ตัวอย่าง ID ขององค์กรหรือหน่วยงานข้างต้น ได้แก่ ธนาคารอิสราเอล (BOI), สำนักงานสถิติแห่งสหภาพยุโรป (ESTAT) และองค์การแรงงานระหว่างประเทศ (ILO) เป็นต้น โดยรายชื่อ agency สามารถตรวจสอบได้ที่

<https://registry.sdmx.org/organisations/agencies.html>

**การกำหนดพื้นฐานของการสร้าง code list** ในการกำหนดพื้นฐานเบื้องต้น code ต้องขึ้นต้นด้วยตัวใหญ่ A-Z 0-9 และขีดล่าง (Underscore) ไม่ให้เน้นตัวอักษรหรือตัวเลขเป็นตัวหนา ซึ่งสัญลักษณ์ขีดล่าง (Underscore) ใช้สำหรับการผสม code แต่ในส่วนชื่อ code name ให้ใช้ 1- 254 ตัวอักษร และตัวอักษรอยู่ในแบบของ UTF-8

ตัวอย่างเช่น \_T หมายถึง ทั้งหมด (Total)

\_U หมายถึง ไม่มีข้อมูล (No data/unknown)

ตัวอย่างการเขียน code

การกำหนด ช่วงชั้นของอายุ (age classes)

โดยกำหนดหน่วยนับของอายุ เช่น

D คือ วัน (day)

W คือ สัปดาห์ (week)

M คือ เดือน (month)

Y คือ ปี (year)

และกำหนดชุดมาตรฐานของตัวดำเนินการ (operation) เช่น

T คือ ใช้กำหนดช่วง (expressing ranges)

\_ คือ ใช้สำหรับเชื่อมรวมระหว่าง code (combination of two code)

X คือ ไม่รวมหรือยกเว้น (except or excluding)

GT คือ มากกว่า (greater than)

LT คือ น้อยกว่า (less than)

GE คือ เท่ากับหรือมากกว่า (equal to or greater than)

LE คือ เท่ากับหรือน้อยกว่า (equal to or less than)

โดยการกำหนดแนวคิด (concept) เช่น

15 – 20 ปีแต่ไม่รวม 16 ปี (from 15 years to 20 years excluding 16 years): Y15T20X16

4 วันหรือมากกว่า (four days or over): D\_GE4

2 และ 3 เดือน (two and three months): M2\_3

## **SDMX GUIDELINES**

# **GUIDELINES FOR THE CREATION AND MANAGEMENT OF SDMX CODE LISTS**

VERSION 3.0

19/1/2018

## Contents

1	<b><i>DOCUMENT HISTORY .....</i></b>	<b><i>1</i></b>
2	<b><i>1. INTRODUCTION.....</i></b>	<b><i>1</i></b>
3	<b><i>2. BASIC PRINCIPLES FOR THE CREATION OF SDMX CODE LISTS .....</i></b>	<b><i>2</i></b>
4	<b><i>3. BASIC CHARACTERISTICS OF SDMX CODE LISTS.....</i></b>	<b><i>3</i></b>
5	<b><i>4. BASIC CRITERIA FOR THE DEVELOPMENT OF SDMX CODE LISTS ....</i></b>	<b><i>4</i></b>
6	<b><i>5. MAINTENANCE OF SDMX CODE LISTS.....</i></b>	<b><i>9</i></b>
7	<b><i>6. OTHER PRACTICAL ISSUES AND RECOMMENDATIONS.....</i></b>	<b><i>9</i></b>

## 8 DOCUMENT HISTORY

Version	Date	Comment
1.0	02/12/2013	Initial version for cross-domain code lists.
2.0	15/01/2015	Adapted to be a guideline for all code lists, not only cross-domain
3.0	19/01/2018	Clarified text on definitions. Revised allowed characters such as leading zeroes in codes (now allowed). Removed superfluous text. Improved examples.

## 9 1. INTRODUCTION

These guidelines are intended to support the creation of SDMX code lists to be used all along the statistical business process and in particular when SDMX is implemented in statistical domains. They are strongly recommended for use when SDMX-compliant data structure definitions (DSDs) are built-up and implemented in statistical domains.

In the [SDMX Checklist for Design Projects](#) and the [modelling guidelines](#), the creation of code lists is done in the sub-process "Fully define code lists".

Originally this document was named "*Guidelines for the Creation and Management of SDMX Cross-Domain Code Lists*". Later, experience showed that these guidelines were also used for the development of other types of SDMX code lists (shared code lists, domain-specific code lists). It was therefore decided to review the document in order to convert it into a guideline applicable to all types of SDMX code lists.

Code lists are created to group related codes<sup>1</sup> in a meaningful, systematic and standard format. They provide lists of codes that objects corresponding to a specific concept can be classified into. Each code should be well described.

Code lists are primarily used to:

- collect, disseminate, exchange and organise information;
- aggregate and disaggregate datasets in a meaningful way for complex analysis;
- present statistical information in a standard way;
- support policy and decision-making;
- standardise the measurement process.

SDMX cross-domain code lists can be found at two distinct places:

1. [SDMX official website](#) (files available in MS-Word format)
2. [SDMX Global Registry](#) (files available in SDMX-ML format)

Other SDMX code lists can be found in regional and other registries.

<sup>1</sup> In the traditional classification sphere, the words "category" or "position" are generally preferred to the word "code"; however, the use of the term "category" could be confusing as it is already used in the SDMX information model within another context.

## 2. BASIC PRINCIPLES FOR THE CREATION OF SDMX CODE LISTS

2.1 SDMX code lists should refer to clear and well-defined statistical concepts, enabling data users to understand the statistical concepts and finally the data sets. Already existing standards (e.g. international classifications) should be taken into account.

2.2 Consistency of the SDMX code lists across statistical domains and over time should be ensured.

### 2.3 Basic definitions

Below are the basic definitions used in this document to describe SDMX code lists.

- A **code list** is a predefined list from which some statistical coded concepts take their values.
    - A **code list identifier** is a unique identifier given to the code list. The code list identifier consists of three mandatory elements: an **id**, a **version number** and a **reference to a maintenance agency**.
    - An **id** is a language-independent set of letters, numbers and/or symbols.
    - A **name** describes the content of the artefact to which the name is attached in a synthetic and clear way. In principle, the default language for names is English (however exceptions are possible, e.g. when geographic entities are expressed in the national languages). Multilingual representations are possible.
    - A **description** allows to describe the content of the artefact to which it is attached in a more detailed fashion than the artefact name. Multilingual representations are possible. In the specific case of code lists, the description is generally used to precisely define the coverage of a code, identifying what is included and what is excluded (e.g. wooden shoes are not considered as shoes but as handicraft).
- A code list may also contain annotations, a uri, a description, attributes indicating the period of validity (e.g. "*valid to*" and "*valid from*") and an attribute indicating whether the code list is final.
- To give SDMX code lists a clear visual identity, the code list identifiers should be prefixed with CL\_.
- A **code** is represented by a mandatory **id** and a mandatory **name**, and an optional **description**. The code may also contain annotations, and a uri<sup>2</sup>.
    - An **id** is a language-independent set of letters, numbers and/or symbols.
    - A **name** describes the content of the artefact to which the name is attached in a synthetic and clear way. In principle, the default language for names is English (however exceptions are possible, e.g. when geographic entities are expressed in the national languages). Multilingual representations are possible.

---

<sup>2</sup> Uri: Uniform Resource Identifier

- A **description** allows to describe the content of the artefact to which it is attached in a more detailed fashion than the artefact name. Multilingual representations are possible. In the specific case of code lists, the description is generally used to precisely define the coverage of a code, identifying what is included and what is excluded (e.g. wooden shoes are not considered as shoes but as handicraft).

While code names and descriptions are meant for interpretation by humans, **ids** are primarily designed to be read by machines. Nevertheless, it can often be helpful for data users that they are meaningful in accordance with the default language used for the name. When choosing the best approach implementers should also consider the possible impact on the code identifiers length.

### Example

Code list	
Code list name	Sex
Code list id	CL_SEX
Code list version	1.0
Code list agency	SDMX
Codes	
Code id	F
Code name	Female
Code id	M
Code name	Male
etc.	

## 3. BASIC CHARACTERISTICS OF SDMX CODE LISTS

Conceptually SDMX code lists can have the following characteristics:

- A code list (e.g. geographical entity) can be referenced by several statistical concepts (i.e. declaring country, country of birth, partner country, etc.).
- The codes used should cover exhaustively the part of reality that is intended to be described by the code list.
- The codes have to be clearly defined. Codes with different coverage must have different code identifiers and names (e.g. Europe including Greenland must have a different identifier and name than Europe excluding it). Different codes should not have the same meaning or coverage.

- The coverage of the codes may however overlap partially within one code list (but may not be identical, except when code lists are based on established standard classifications where such repetitions are common, see point 6.1.2). This means that the content of the categories is not necessarily mutually exclusive as is the rule for established statistical classifications.
- Codes may be at different levels of granularity.
- Multiple hierarchies (hierarchical code lists<sup>3</sup>) can be defined on top of a flat code list. An SDMX code list should be flexible in terms of allowing different possible hierarchies and be extendible by additional codes that may disaggregate or aggregate codes that are already in the list as well as by codes that extend the coverage of the code list. Each possible hierarchy may use all codes from the flat code list or just a subset. A flat code list provides the reservoir of codes for the hierarchies.
- New codes or levels need to be accommodated if needed (i.e. if a new country is recognised as a sovereign state the geographical entity code list should have a new version including it).
- Re-use of code identifiers between versions of the same code list should be avoided. In ISO-2 country code CS - Czechoslovakia was suppressed in 1993 and reused as CS - Serbia and Montenegro in 2003. This change was/is extremely difficult to handle, as many data systems will contain historical data for Czechoslovakia.

## 4. BASIC CRITERIA FOR THE DEVELOPMENT OF SDMX CODE LISTS

4.1 The following basic criteria should be respected when defining SDMX code lists:

- It is highly recommended that code lists be consistent, to the largest extent possible, with internationally agreed standards, whenever they exist, e.g. International Organization for Standardization (ISO), United Nations and other international organisations code lists. It is of no use creating a new code list where one already exists. The following order of priority is suggested when considering the use of existing code lists:
  - international standard classifications or code lists;
  - international classifications or code lists supplemented by other international and/or regional institutions;
  - standardised classifications or code lists used by individual international institutions.
- If official classifications are used for defining the SDMX code lists, totals, aggregates and other additional codes should be added following the recommendations of the issuing organisation for adding new elements (e.g. the addition of the world regions EU27, EFTA, etc to the ISO 3166 alpha 2 country list);

<sup>3</sup> The "Hierarchical code list" construct used in SDMX should not be mixed up with the concept used in traditional statistical classifications. In the latter case, the codes are organised based on one strictly defined hierarchy only. In the former case, several hierarchies can be defined. An example of this is a list of codes that support the theme of geographical location. Such a hierarchical code list could be viewed according to many different hierarchies. A political hierarchy would comprise an administrative regional breakdown within a country, a geographical breakdown would comprise a placing of the countries in continents, and an economic breakdown might place the countries in one or more economic communities (e.g. many of the countries in "Europe" could be both a part of the European Union and the OECD communities).



- When designing code lists, the needs of all phases of the statistical business process and of all statistical domains using the respective SDMX code lists should be taken into consideration. However due to legacy systems, it may not always be possible to accommodate all needs from the start (e.g. different habits in data exchange and in dissemination). In such cases, some transition period will often be necessary before reaching convergence.

4.2 Technically, SDMX code lists should respect the principles described in the following paragraphs.

- SDMX code ids take values from uppercase A to Z, 0 to 9 and "\_" only<sup>4</sup>. No other characters should be used.
- Even though technically allowed in the standard, it is highly recommended not to use lower case characters in order to avoid possible confusion and technical issues with upper case characters. Exceptions should only be made for code lists based on classifications managed by external bodies (such as ISO). In case an external classification is using lower case characters, they may also be used in the respective SDMX code lists in order to make their usage more intuitive for regular users of the classification. However, it is still not allowed to create two code list items that only differ by the case of the characters (e.g., "EN" and "en"). If for whatever reason an organisation opts for lower case characters, all coded information should then be in lower case characters.
- Accented characters for a code are not allowed by the standard.
- Underscore ("\_") is generally used for the combination of codes (whether consecutive or not).
- A list of generic codes to be used for all SDMX code lists has been developed. This generic list includes concepts which can be expected to appear in many, if not, all code lists, e.g. "total", "non-response", "not allocated", "unknown, other, etc.

4.2.1 Code names should be between 1 and 254 characters. The characters used should belong to the UTF-8 character set.

- Code names are in general defined in English; other language versions may be added. Certain exceptions can be considered in the case of codes which are normally not translated (e.g. regions or territories for which the national designation could be used, agencies for which a national and English label could be used, etc.).

4.2.2 Descriptions could be used if more details on the contents or on the code descriptions are needed. Multilingual representations are possible.

#### 4.2.3 Meaningful code identifiers vs meaningless code identifiers

Code identifiers can be either meaningful, i.e. the identifier is a shortened representation or expression of the code name (e.g. "DE\_FIN" for "German Ministry of Finance"), or meaningless, i.e. the code identifier is chosen at random and conveys no information whatsoever as to the content of the code name (e.g. "6E" standing for "European Space Agency").

<sup>4</sup> It should be noted that the list of allowed characters has been restricted compared to the general provisions of the SDMX technical standards. The SDMX schema defines that valid characters for Code are [A-z 0-9 \_@\$%-]. For code lists, it was felt that allowing some of these characters might generate confusion (e.g. mixing lowercase and uppercase might give the feeling that each of them has a distinct meaning) or potential problems (e.g. in validation procedures where the dash ("-") could be considered as an operator for subtraction).

In SDMX, both coding systems are acceptable, although it is recommended to use meaningful codes if possible. It is up to implementers to decide which coding system they want to use, considering that there could be a trade-off between meaningfulness of identifiers and their length which impacts the maintainability of identifiers/key series codes, size and costs of data files and databases. The choice of the coding system may be influenced by parameters like the number of identifiers to code, the complexity of the concepts to code, etc.

In cases where self-explanatory (= meaningful) identifiers are chosen, implementers should pay attention not to use excessively long codes.

Meaningful codes can be meaningful in one language only; by default, this language is English.

#### 4.2.4 Expressing numeric ranges in code identifiers

SDMX implementers may wish to express numeric ranges in code identifiers. This is allowed but in no way imposed.

As example, let us take a code list dealing with age classes. The basic units used for representing age are "Y" (Year), "M" (Month), "W" (Week) and "D" (Day). To express concepts like "from 15 years to 20 years excluding 16 years", "over 30 years", "less than 50 years", "two and three months", "four days or over", "three years or less", the following set of standard "operators" is proposed:

- "T" for expressing ranges (from 3 to 9),
- "\_" for the combination of two codes, whether consecutive or not (5 and 6; A and F),
- "X" for expressing "except" or "excluding",
- "GT" for "greater than", "LT" for "less than", "GE" for "equal to or greater than" and "LE" for "equal to or less than".

The result for the examples mentioned in the previous paragraph is as follows:

- from 15 years to 20 years excluding 16 years: Y15T20X16
- over 30 years: Y\_GT30 (in this case the "\_" is used to make a clear distinction between the unit and the "operator" "GT").
- less than 50 years: Y\_LT50
- two and three months: M2\_3
- four days or over: D\_GE4
- three years or less: Y\_LE3

Such a coding system has limits in what it can do. It will most probably not be appropriate for coding ranges with decimals (ending up in too long code identifiers) or too complex concepts. Again, it will be up to implementers to decide whether this coding system can meet their specific needs.

#### 4.2.5 Generic codes

These codes are recommended to be used when compiling code lists that require the items in the table below, e.g. when a "Not applicable" item is required. The codes provide code values for concepts referenced in a large number of code lists.

The codes proposed here are present in a very large number of SDMX code lists because they cover very general and extensively used concepts. Thus, the main purpose of this list of codes is to propose standardised code identifiers which can be extensively reused.

213 The leading underscore is used to avoid clashes with other codes and to singularise these  
214 concepts.

215 These codes should be considered as "reserved codes" or "protected codes", meaning that It  
216 is strongly recommended not to use them for other purposes than the ones described below.

217 The codes can also be used as suffixes or be augmented with sequential numbers, especially  
218 if a category needs to be used multiple times. For example, multiple sub-totals  
219 (e.g. CATTLE\_S1, CATTLE\_S2, PLANTS\_S1) or multiple residual categories  
220 (e.g. CATTLE\_O, PLANTS\_O) may be required.

221 In cases where the use of the suffixes below could lead to confusions and misunderstandings,  
222 implementers may of course implement other solutions. Let us take the example of a code list  
223 where letters are used as codes (e.g. the United Nations ISIC classification): in this case, code  
224 element M\_O could be interpreted either as "sections M and O" or as "section M not elsewhere  
225 classified". In this case implementers could create codes M\_OTH or M\_OTHER to differentiate  
226 the suffixes.

Recommended Code Value	Recommended Code Description	Annotation
_L	Local extension	To be used in order to uniquely identify local (e.g. national, regional, sub-regional) extensions of SDMX code lists and by doing so: <ul style="list-style-type: none"> <li>• avoid that local extensions conflict with existing codes, and</li> <li>• avoid that revisions of the underlying code lists conflict with the local extensions in terms of coding.</li> </ul>
_N	Non response	Failure to obtain a measurement on one or more study variables for one or more elements in a survey.
_O	Other	Used to cover residual information not contained in other categories of the code list (in some contexts, e.g. classifications, referred to as n.e.s., not elsewhere specified, n.e.c., not elsewhere classified, etc.)
_S	Subtotal	Used for expressing intermediate totals
_T	Total	Used for expressing totals
_U	No data/unknown	Failure to obtain a measurement (e.g. non response, no data available, information not known by the respondent unit, etc.)
_X	Not allocated/unspecified	Used where the value for a particular variable falls outside the expected range. An example could be the failure to allocate a classification to a particular unit due to insufficient information, and/or if further breakdown over any related items mentioned in code list not available

Recommended Code Value	Recommended Code Description	Annotation
<b>_Z</b>	Not applicable	<p>Used in cases where the coding of a concept is technically required (dimension or mandatory attribute), but does not have a statistical meaning for a specific series or observation.</p> <p>Examples of relevant usages of <b>_Z</b> are:</p> <ul style="list-style-type: none"> <li>• In a survey that has questions on “Marital Status” and “Sex of Partner”, for a response of Marital Status: Single, the relevant response for Sex of Partner is “Not applicable” (<b>_Z</b>)</li> <li>• In a labour statistics reporting framework that has a dimension “Outside Labour Force Reason”. If the observation entity is inside of the labour force, the response for Outside Labour Force Reason is “Not applicable” (<b>_Z</b>)</li> <li>• In macro-economic statistics, a DSD may contain a dimension for breakdowns of economic activity. Some series cannot be broken down by economic activity (e.g. statistical discrepancy between GDP approaches, certain taxes and subsidies, producer price index). They could be coded as “Not applicable” (<b>_Z</b>) for the activity dimension in the data message.</li> </ul> <p>The code <b>_Z</b> should be used sparsely<sup>5</sup>, since it may complicate validation and transformation formulas or pivot visualisations based on SDMX messages. The code <b>_T</b> (Total) should be used instead in cases where the applicability of the concept is not entirely clear, and where the concept may be relevant only under certain circumstances, or where the applicability of a concept for a certain series or dataflow depends on the context (e.g. a tax breakdown that exists in one country but not in another) in order to simplify the coding across data flows and countries.</p>

<sup>5</sup> At DSD design time the pros and cons of using **\_Z** should be considered carefully. The occurrences of **\_Z** can be reduced by:

- using the semantically approximate **\_T** instead of **\_Z**;
- having a larger number of compact DSDs to avoid the need to introduce **\_Z**;
- concepts with **\_Z** could be defined as attributes rather than dimensions;
- a dimension containing **\_Z** could be merged with another (conceptually related) dimension as a sub-hierarchy.

## 5. MAINTENANCE OF SDMX CODE LISTS

### 5.1 Maintenance agency

Within SDMX, one or several maintenance agencies for each code list or family of code lists has/have to be identified<sup>6</sup>. For SDMX cross-domain code lists, the maintenance agency (marked as "SDMX") is the SDMX Statistical Working Group (SWG). For other code lists used in international data exchange this is - in general - one or several international organisation(s) linked to SDMX. Clear rules for the maintenance of the SDMX code lists need to be established. Changes to a given code list will lead to a new version of the code list. SDMX code lists are stored in the SDMX Global Registry<sup>7</sup> for making them accessible to SDMX implementers. Other code lists are stored in regional or other *ad hoc* registries.

### 5.2 Versioning of code lists

A general document providing guidelines on the versioning of SDMX artefacts can be found on the SDMX official website<sup>8</sup>.

### 5.3 No retroactivity in case of implementation of new code lists

The adoption of a new code list has no retroactive effect on existing code lists. Therefore, the implementation of a new code list will not require from SDMX implementers that they revise existing code lists.

## 6. OTHER PRACTICAL ISSUES AND RECOMMENDATIONS

### 6.1 Treatment of traditional statistical classifications

This section addresses the specific issues linked to established statistical classifications, i.e. the main hierarchical classifications used in the statistical sphere (e.g. United Nations ISIC<sup>9</sup>, European Union NACE<sup>10</sup>, North American NAICS<sup>11</sup>, Australian/New Zealand ANZSIC<sup>12</sup>, etc.).

The recommendations presented in this section apply exclusively to this special type of classifications and should not be taken into account for the construction of other SDMX code lists.

#### 6.1.1 Concepts referred to by multiple code lists

Some SDMX concepts (e.g. ACTIVITY) may have global, regional and/or national classifications and variants (different versions of the same classification). This means that a DSD may need to make reference to multiple different representations of the concept of activity, namely a global one, a regional one and a national one.

<sup>6</sup> Although in the SDMX registry only one agency is allowed for the maintenance of a code list, entries can be created for accommodating 2 or more agencies (e.g. ECB/ESTAT for the European Central Bank and Eurostat).

<sup>7</sup> <https://registry.sdmx.org/FusionRegistry/>

<sup>8</sup> Guidelines on the Versioning of SDMX Artefacts: [https://sdmx.org/?page\\_id=4345#Versioning](https://sdmx.org/?page_id=4345#Versioning)

<sup>9</sup> International Standard Industrial Classification of All Economic Activities.

<sup>10</sup> Statistical Classification of Economic Activities in the European Community.

<sup>11</sup> North American Industry Classification System, version used in the United States of America

<sup>12</sup> Australian and New Zealand Standard Industrial Classification

258 There are two options for dealing with this issue :

- 259       • Create a single Codelist containing the required variants of the classification using  
260       prefixes to distinguish between the variants, e.g. in the ISIC classification version 3 for  
261       Transportation and storage is ISIC4\_H. In NACE revision 2 the same item is  
262       NACE2\_H.<sup>13</sup>
- 263       • Create distinct DSDs which each have one "activity" concept but with a different code  
264       list attached.

---

<sup>13</sup> Ideally, there should be a mapping created between the prefixed variant codes and the original classification codes, and made available in a SDMX public registry. This would avoid redoing the same mapping work by different institutions.