Structural metadata

Mr. Andrea Bruni

ISTAT | Directorate for External Relations, International Affairs, Press Office and National Statistical System Coordination

Outline

Classifications and Code lists

GSIM and SDMX

Strategy

Hands-on

Istat

Classifications and Code lists

Istat

Recalling the definition

Used to identify data:
consist of identifiers and descriptors that are essential for discovering, organizing, retrieving and processing a statistical data set

Examples: titles, subtitles, short descriptions, dimension names, variable names, dictionaries, dataset technical descriptions, dataset locations, keywords for finding data, units of measurement (e.g., EUR), code lists (e.g., for territorial coding), data formats, potential value ranges, time dimensions, value ranges of flags, classifications used, etc.

Data and structural metadata MUST go together.

Structural metadata harmonisation

The compilation of statistics for a specific domain is usually considered, by those who do it, a highly specialized activity. As a result, the production process is separated according to the statistical domain.
This means lack of harmonization with respect to how data is organized, what metadata is provided with and how it is exchanged.
Even within a NSI it can be difficult to share, for example, IT applications across domains, thus creating the potential for inefficiency and duplication of effort.

For the end users, it is often difficult to use statistical information on different subjects or from different providers in an efficient way.

The harmonisation of variables and **their (coded) representation** is an essential condition for generating efficiency gains in:
- data collection (e.g., streamlining of surveys)
- data processing (e.g., record linking, data imputation, data matching)
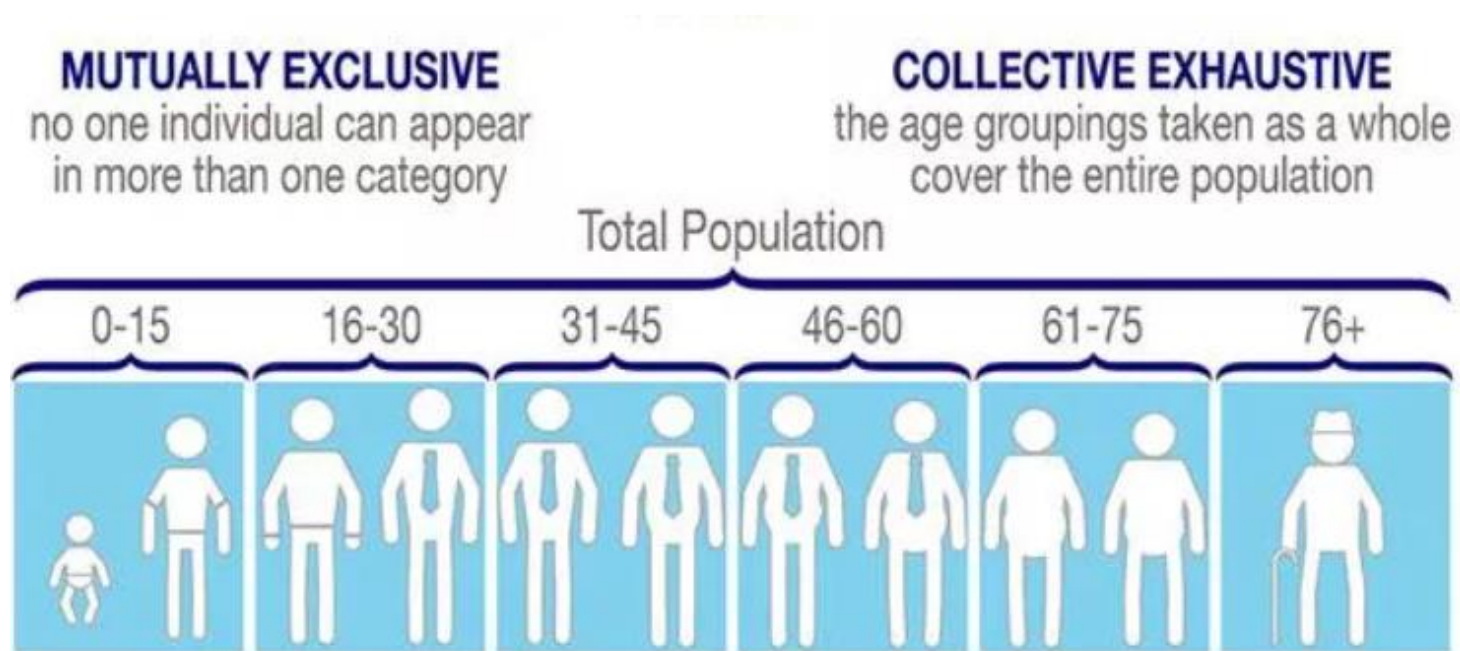- dissemination (e.g., improved capacity to cross-tabulate data)

Key concepts

Integrating different sources means understanding the metadata associated with different data sets, often using different terms with the same meaning or same terms with different meaning.

- **harmonization**: everyone speaks a common language, independently on the statistical theme

- metadata **reuse**: processes should draw pre-existing terms

- **traceability**: every step of the data production process is operationally defined (transparency, automation)

- **search** functionalities (external users formulate a question of information which is always linked to a metadata)

Istat

Classification

A classification is a "SET OF DISCRETE, **MUTUALLY EXCLUSIVE** AND **COLLECTIVE EXHAUSTIVE** CATEGORIES"

no overlapping & no missing

MUTUALLY EXCLUSIVE
no one individual can appear in more than one category

COLLECTIVE EXHAUSTIVE
the age groupings taken as a whole cover the entire population

Total Population

| 0-15 | 16-30 | 31-45 | 46-60 | 61-75 | 76+ |

Checklist for developing a statistical classification

Best_practices_for_Classif_UN_Expert_Group

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

The essential components of a statistical classification are:
- a consistent conceptual basis;
- a flat or hierarchic structure;
- categories that are mutually exclusive and exhaustive;
- definitions that are clear and unambiguous, and which define the content of each category;
- that it is up-to-date and relevant;
- that it is robust enough to last for a period of time;
- that it meets user needs;
- that it provides comparability over time and between collections;
- that it provides guidelines for coding and output of data collected using it;

Istat

Essential components of a statistical classification

Best_practices_for_Classif_UN_Expert_Group

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

1. Status of the classification
2. Which data collections will use this classification?
3. What are the underlying concepts used in this classification?
4. Scope of the classification
5. Primary uses of the classification
6. User Consultation
7. What are the classification criteria?
8. Structure of the classification
9. Are the proposed categories well defined?
10. Appropriateness of code structure
11. Relationship with other classifications
12. Statistical Balance

Code lists

Code lists are created to group related items in a meaningful, systematic and standard format. They provide lists of codes to classify according to a specific concept (e.g., sex, age).

Actually, **a code list is just an object containing a list of codes**…
a set of values to be used in the representation of a Concept (Dimension or Attribute) in Data/Metadata Structure Definitions.

Code lists are primarily used to:
➢ Collect, disseminate, exchange statistical information;
➢ Aggregate and disaggregate datasets in a meaningful way for complex analysis;

Standard code lists should be used all along the statistical business process, (i.e., design, collection, aggregation, dissemination and archiving)

Istat

Code lists vs Classification

A code list should 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 (e.g., totals, aggregates).

The coverage of the codes may however overlap partially within one code list (but may not be identical). This means that the content is not necessarily mutually exclusive as is the rule for established statistical classifications.

Concept: CIVIL STATUS OF AN

INDIVIDUAL

Codelist: CL_CIVIL_STATUS*

S Single

M Married

W Widowed

D Divorced

M_W_D Married at least once

W_D Once married

_T Total

* "registered partnership" and "legally separated" items not considered
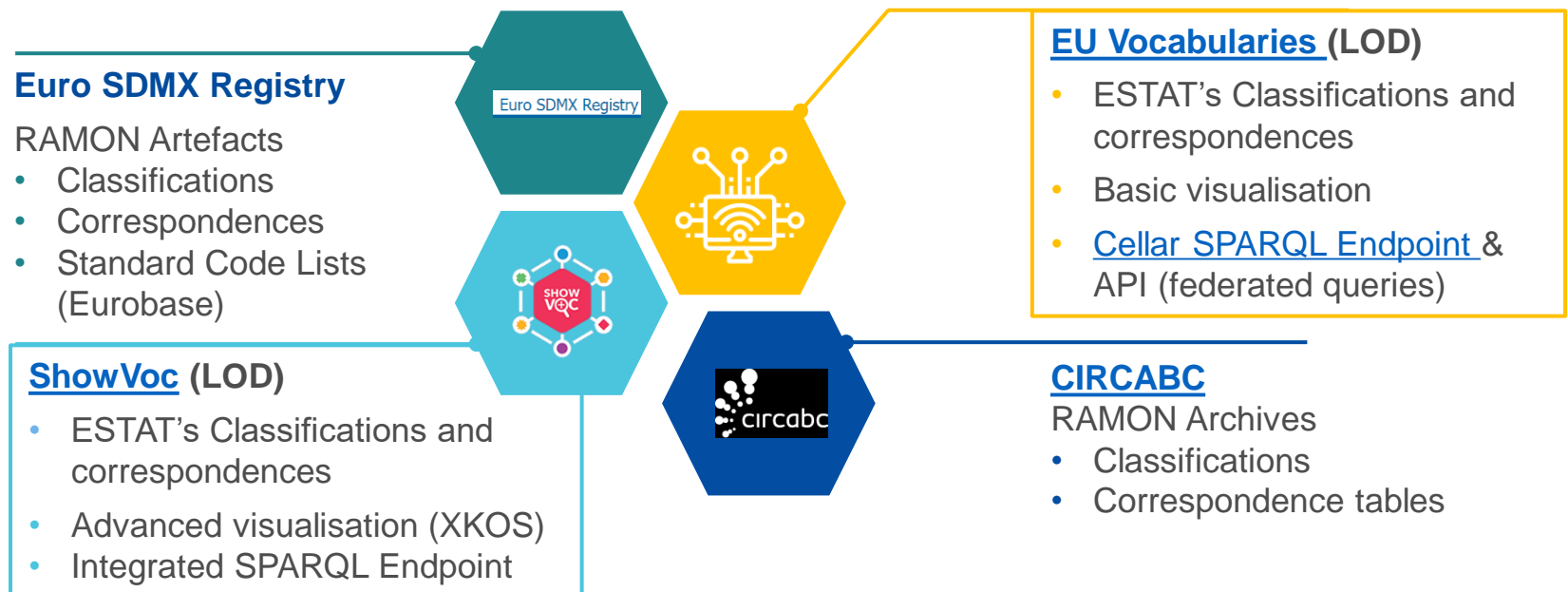
11

Istat

Code lists standardisation

Code lists not harmonised means that different codes are used for the same statistical concept (e.g., for Manufacturing industry the codes "RD", "B0200", "SE0_4" and "TOT_MANUF" are used in four different production databases, while the standard code for this NACE section is "D" in the Reference).

This situation leads to extra work and is a permanent source of errors.

The use of **standard code lists** will help to work even more efficiently, easing the maintenance of and reducing the need for mapping systems and interfaces delivering data/metadata to them.

Istat

ShowVoc Webinar – October 2023

Eurostat metadata dissemination – Platforms (replacing RAMON)

Euro SDMX Registry

RAMON Artefacts
- Classifications
- Correspondences
- Standard Code Lists (Eurobase)

ShowVoc (LOD)

- ESTAT's Classifications and correspondences
- Advanced visualisation (XKOS)
- Integrated SPARQL Endpoint

EU Vocabularies (LOD)

- ESTAT's Classifications and correspondences
- Basic visualisation
- Cellar SPARQL Endpoint & API (federated queries)

CIRCABC

RAMON Archives
- Classifications
- Correspondence tables

Euro SDMX Registry

circabc

European Commission

ShowVoc Webinar – October 2023

Other dissemination platforms

	EU Vocabularies	Euro SDMX Registry	CIRCABC (Archives)
Statistical Classifications	✓	✓	✓
Correspondence tables	✓	✓	✓
CODED		✓	
Metadata catalogues	✓	✓	
Standard Code Lists		✓	

https://ec.europa.eu/eurostat/web/metadata
https://ec.europa.eu/eurostat/web/metadata/classifications

European Commission

International standards

UNECE Standard

The Generic Statistical Information Model (GSIM) models all of the information needed for the statistical production process

At Statistics Flanders

Vlaamse overheid

(3 workshops) → distilled to some 40 objects

Istat

GSIM benefits

Immediate benefits

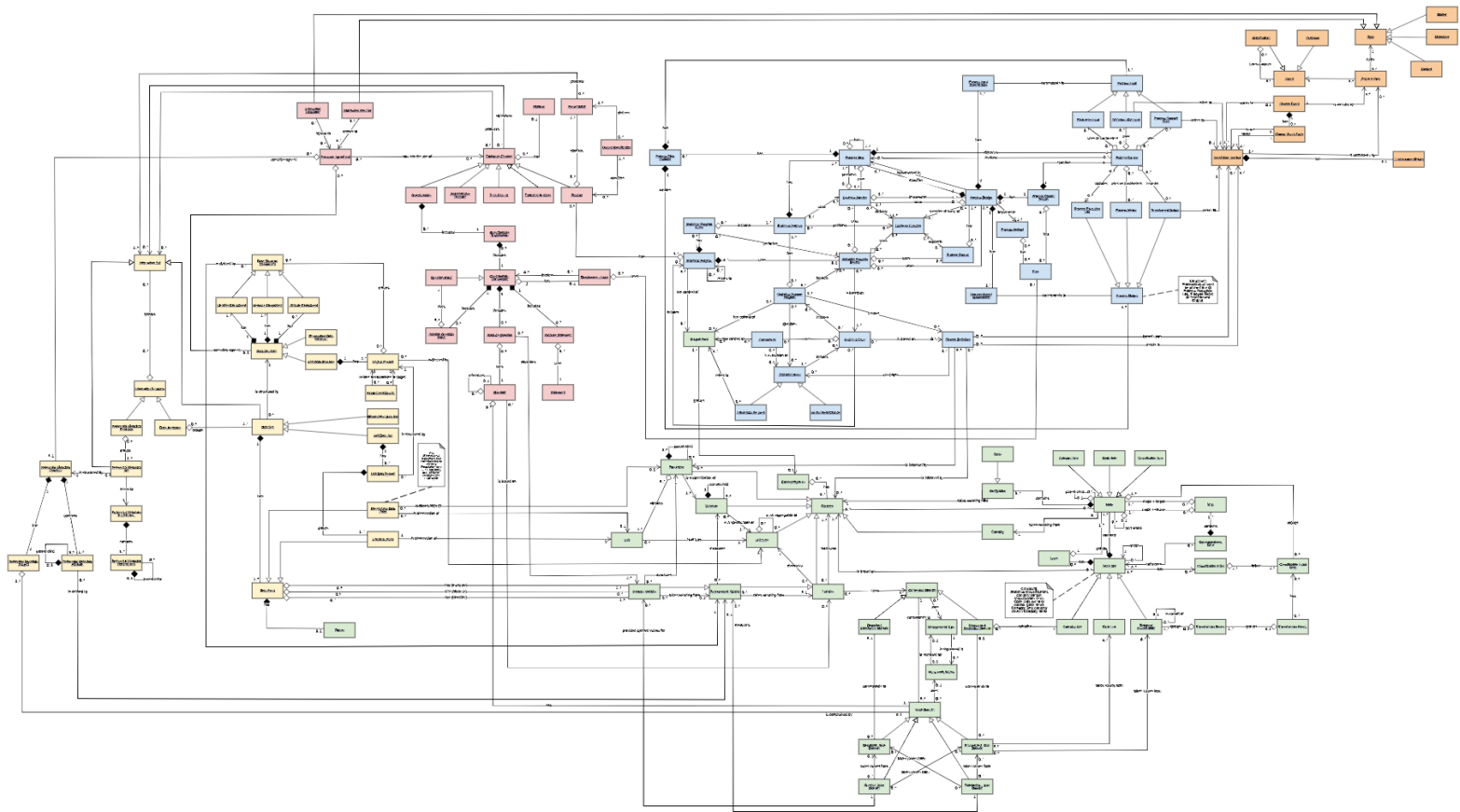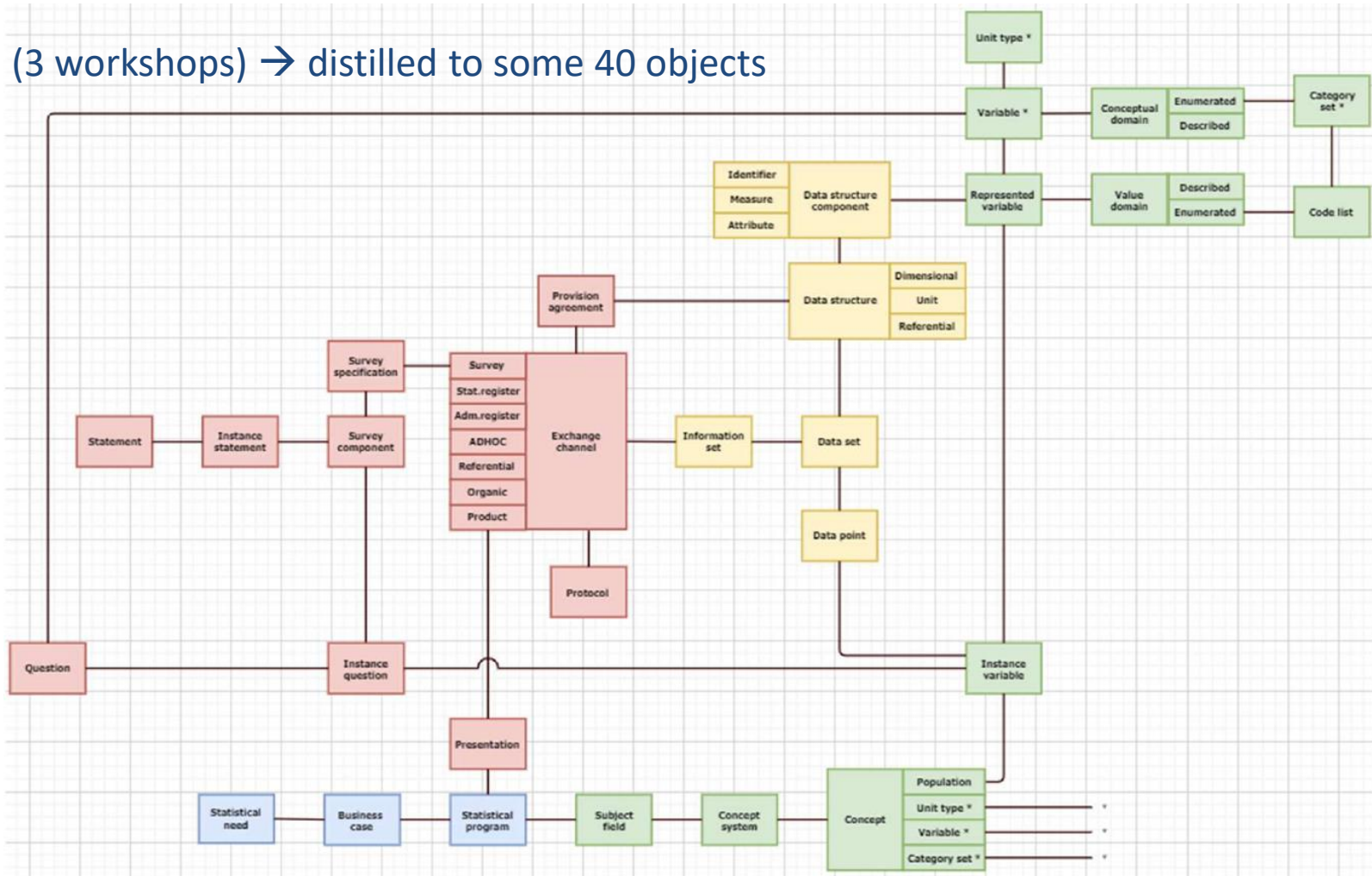We can **talk** about the same information objects **with the same names**.

GSIM provides a common language to improve communication at different levels:

- Between the different roles in statistical production (business and IT experts);
- Between the different statistical subject matter domains;
- Between statistical organizations at national and international levels

Long term benefits

Reuse and sharing of methods, components and processes: the use of GSIM will reduce workloads as many processes can be repurposed and reused.

Statisticians less reliant on information technologists.
In the "stove-pipe" approach to statistical production, subject matter is heavily dependent upon the ITs in the design, build and production of statistical systems.

An increase in the use of standardized applications, which can easily be shared across domains, will enable statisticians to more easily work in different domains.

Istat
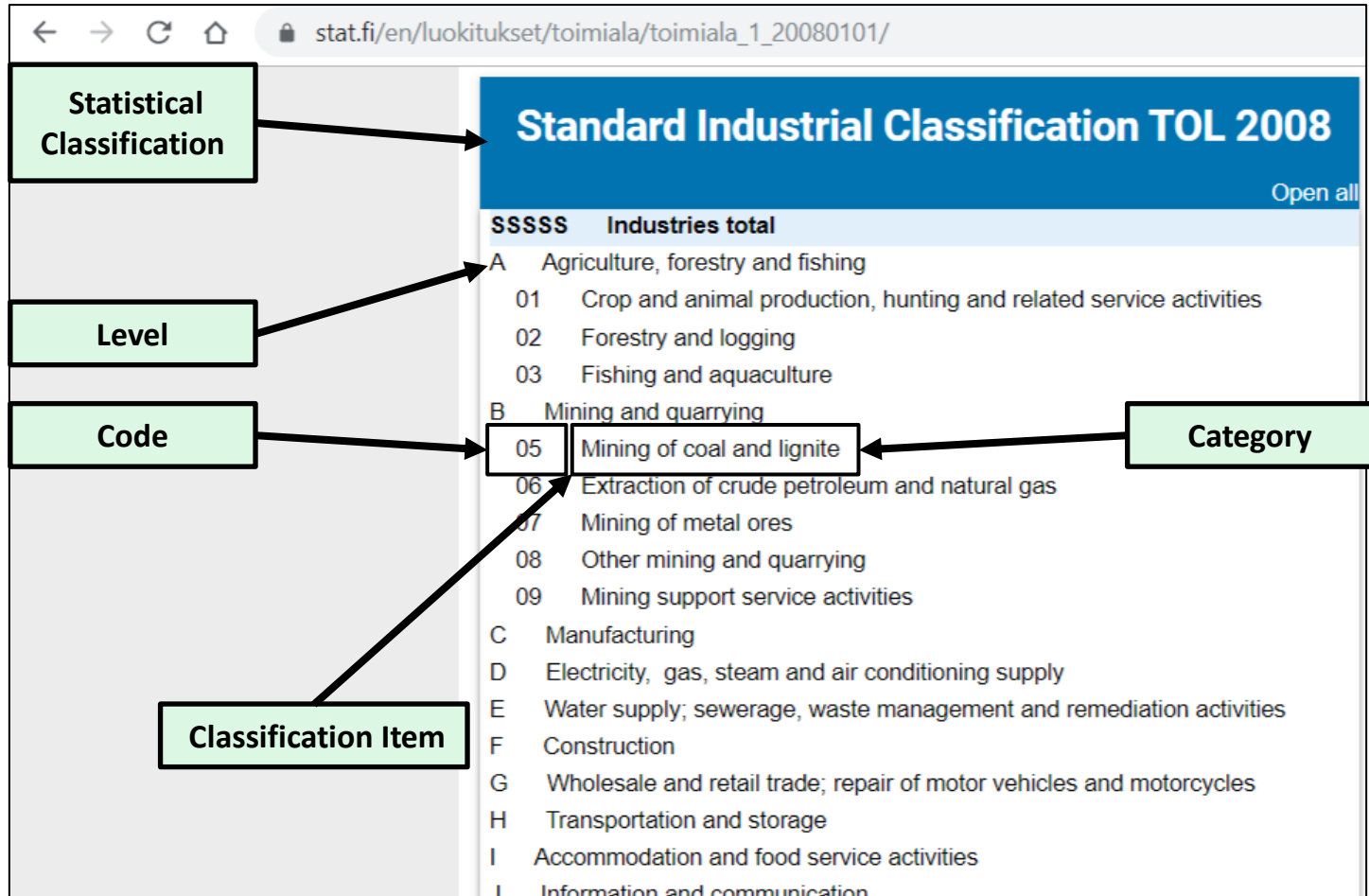
Structure objects

e-Training ▶ YouTube
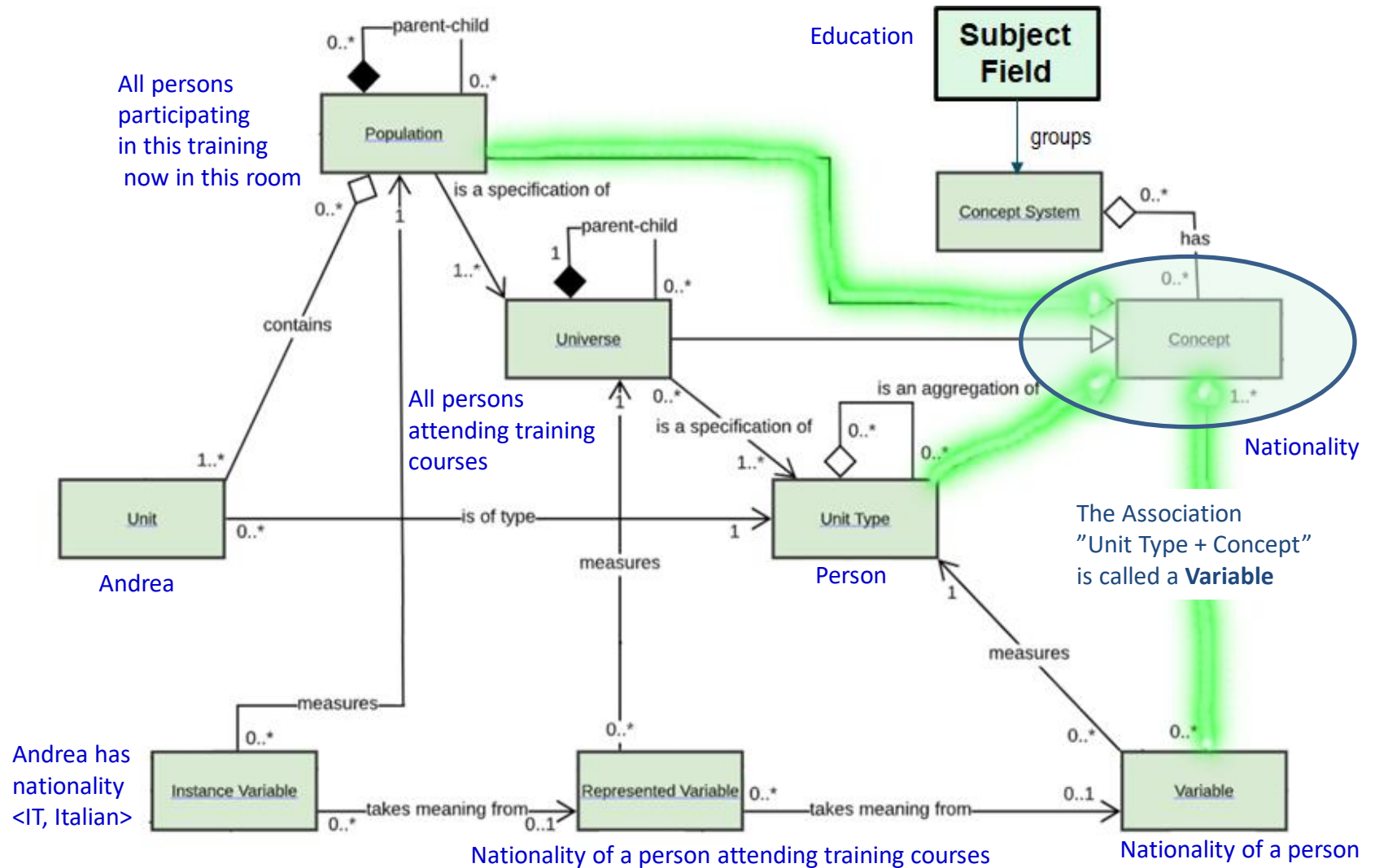
It is too restrictive to say a *Classification Series* (CS) is used to classify one and only one *Unit Type* (UT). An example is a standard occupational classification (SOC). A SOC could be used to classify people based on the jobs they hold, employers based on each of the jobs they fill, or job categories managed by an HR department based on necessary skills and duties.

Each CS classifies one or more UT, and each UT is classified by zero or more CS.
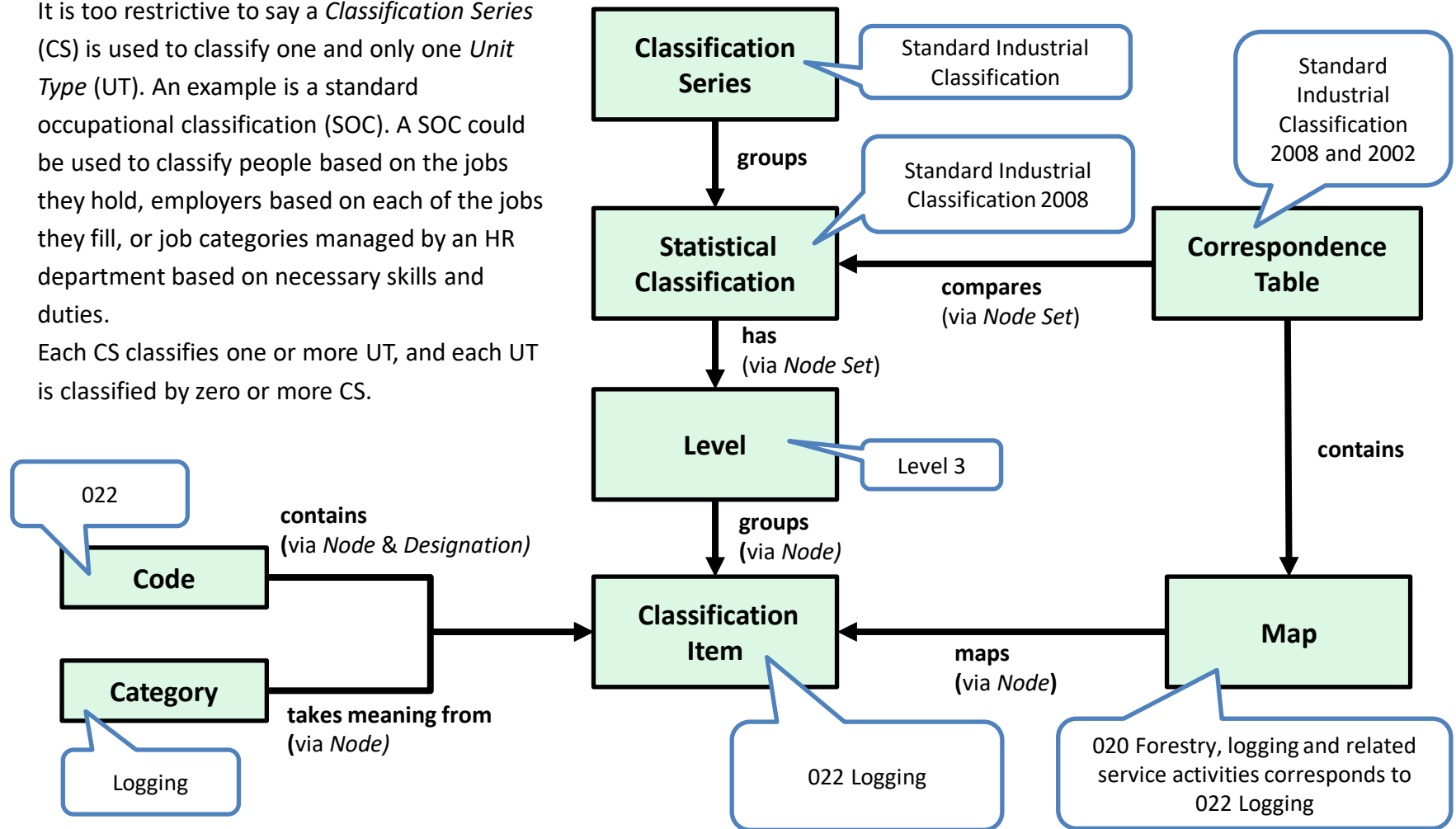
Classification Series

Standard Industrial Classification

Standard Industrial Classification 2008 and 2002

groups

Standard Industrial Classification 2008

Statistical Classification

Correspondence Table

compares (via *Node Set*)

has (via *Node Set*)

contains

Level

Level 3

022

contains (via *Node* & *Designation)*

groups (via *Node)*

Code

Classification Item

Map

maps (via *Node)*

Category

takes meaning from (via *Node)*

Logging

022 Logging

020 Forestry, logging and related service activities corresponds to 022 Logging

Istat

e-Training ▶ YouTube

Unit Type: Households; *Population*: Households resident in Italy in 2016
Represented Variables: Annual average household income, Territory,…

annual average households income						
	Select time	2016				
	Number of components	one	two	three	four	five or more
		▲ ▼	▲ ▼	▲ ▼	▲ ▼	▲ ▼
Territory	**Households main income source**					
■ Italy	employee income	21 628	33 540	38 293	43 195	43 952
	self-employed income	20 518	39 951	43 003	41 482	43 587
	public transfers income	16 958	28 793	39 149	36 550	32 825
	other type	11 224	27 246	26 584 (n)	19 828 (0)	..
	total	18 257	31 075	38 765	41 693	41 878
Nord-ovest		19 885	33 626	44 009	50 087	46 954

identifiers   measure   attribute

Istat

GSIM and SDMX

Measure: The role given to a *Represented Variable* in the context of a *Data Structure* to hold the underlined observed/derived values. For example, Economic Activity in a *Unit Data Set* (microdata) or Number of citizens in *Dimensional Data Set* (aggregate data)

is defined by → **Represented Variable**

e.g., Territory, Economic activity

Dimensional Data Structure Component	Identifier
	Measure
	Attribute

GSIM

is defined by →

DSD (Data Structure Definition)	Dimension
	Measure
	Attribute

Concept ← groups ← **Concept Scheme**

e.g., Territory, Economic activity

sdmx

Istat
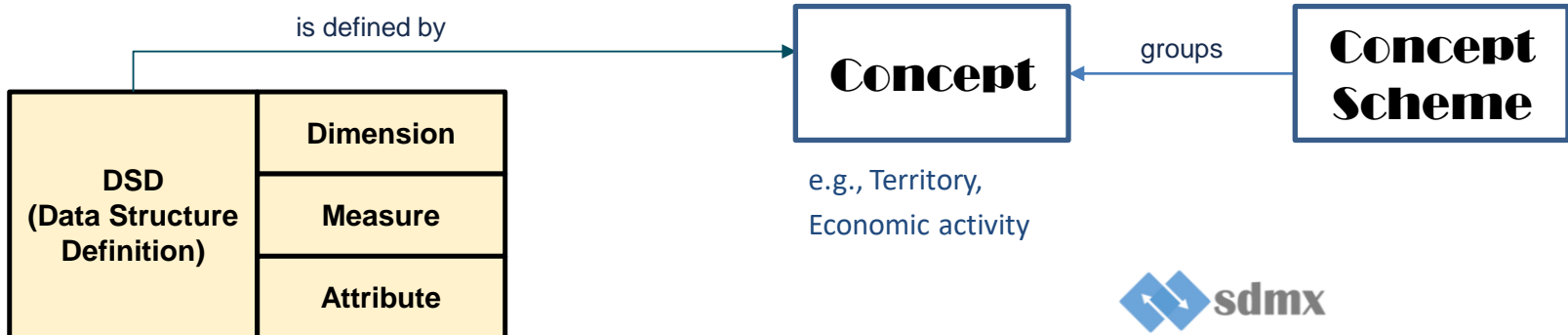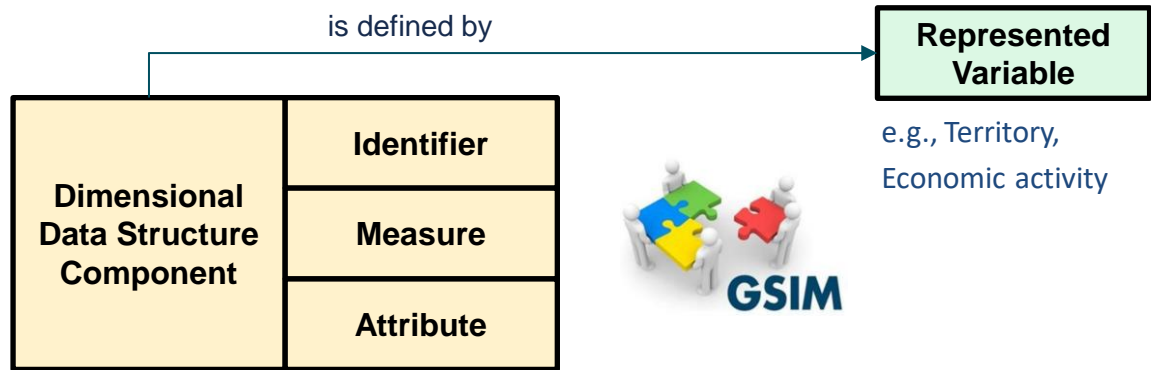
GSIM and SDMX

SDMX has been a real success for the harmonization of data/metadata exchange GSIM could help a lot a statistician in using SDMX, in properly modelling a data cube, by assigning a concrete role to concepts before their use in a SDMX Data Structure Definition (DSD).

SDMX_Glossary_Version_2_1_December_2020.htm
Type = Cross-domain concept

Concept Scheme(s)

Codelists

DSD

Istat

Analysis

Identify the CONCEPTS

Cross-domain concepts

Seasonal adjustment

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

Context   Seasonal adjustment removes the effects of events that follow a more or less regular pattern each year. These adjustments make it easier to observe the cyclical and other non-seasonal movements in a data series.

Type   Cross-domain concept

Concept ID   SEASONAL_ADJUST

Recommended representation   Codelist

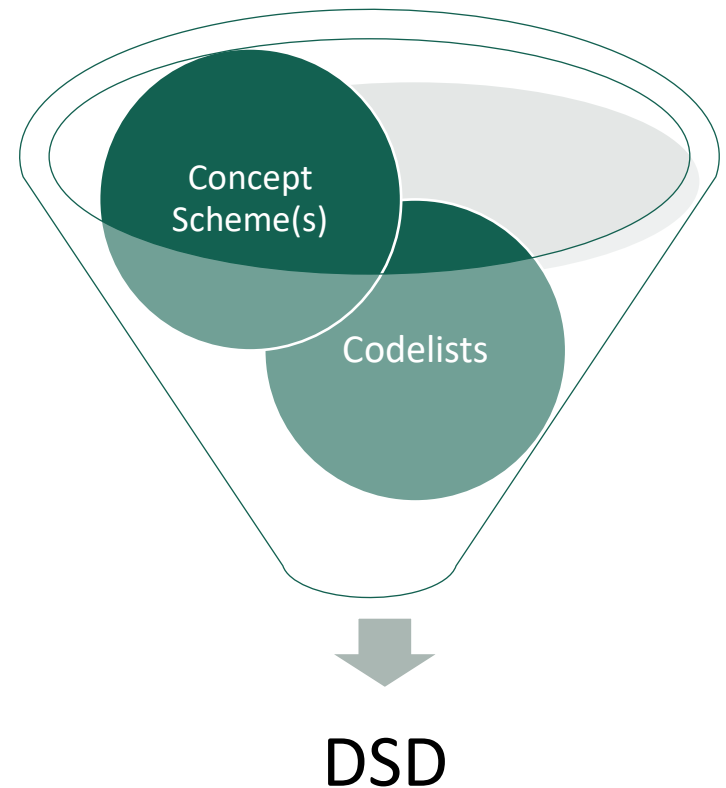**Codelist ID**   CL_SEASONAL_ADJUST

Related terms   Adjustment

                      Price adjustment

Source   Australian Bureau of Statistics, "An Analytical Framework for Price Indexes in Australia: Glossary and References", Ca
(http://www.abs.gov.au/ausstats/abs@.nsf/bb8db737e2af84b8ca2571780015701e/ff4de83064a2e425ca25697e0018fd44!OpenDocument)

Other link(s)   Codelist CL_SEASONAL_ADJUST (https://sdmx.org/?page_id=3215)

                      U.S. Bureau of Labor Statistics, Online glossary, last consulted August 2020 (http://www.bls.gov/bls/glossary.htm)

Reference area

Definition   Country or geographic area to which the measured statistical phenon

Context   The concept refers to the country, geographical or political group of c

                The concept is subject to a variety of hierarchies, as countries compr
statistical data are produced internationally on a separate and independent basis.

Type   Cross-domain concept

Concept ID   REF_AREA

Recommended representation   Codelist

Codelist ID   CL_AREA

Related terms   Counterpart reference area

Geographical coverage

Source   SDMX, "Metadata Common Vocabulary", 2009 (https://sdmx.org/wp

Other link(s)   Codelist CL_AREA (https://sdmx.org/?page_id=3215, see under "Ge

Economic activity

Definition   Combination of actions that result in the production, distribution and consumption of goods or services.

Context   An activity can be said to take place when resources such as equipment, labour, manufacturing techniques or products are com
resources, a production process and an output of products.

Type   Cross-domain concept

Concept ID   ACTIVITY

Recommended representation   Codelist

Codelist ID   CL_ACTIVITY

Related terms   Economic sector

Source   SDMX, "SDMX Glossary Version 1.0", February 2016 (https://sdmx.org/wp-content/uploads/SDMX_Glossary_Version_1_0_February_2016.docx)

Other link(s)   Codelist CL_ACTIVITY (https://sdmx.org/?page_id=3215)

Cross-domain codelists

These are the cross-domain code lists that have been officially adopted so far.

- Activity (CL_ACTIVITY)
- Age (CL_AGE)
- Break Reason (CL_BREAK_REASON)
- Civil (or marital) status (CL_CIVIL_STATUS)
- Classification of Individual Consumption According to Purpose (COICOP) (CL_COICOP)
- Classification of the Functions of Government (COFOG) (CL_COFOG)
- Classification of the Outlays of Producers According to Purpose (COPP) (CL_COPP)
- Classification of the Purposes of Non-Profit Institutions Serving Households (COPNI) (CL_COPNI)
- Confidentiality status (CL_CONF_STATUS)
- Currency (CL_CURRENCY)
- Decimals (CL_DECIMALS)
- Degree of Urbanisation (CL_DEG_URB)
- Frequency (CL_FREQ)
- Geographical areas (CL_AREA)
- Observation status (CL_OBS_STATUS)
- Occupation (CL_OCCUPATION)
- Organisation concepts (CL_ORGANISATION)
- Reliability (CL_RELIABILITY)
- Seasonal adjustment (CL_SEASONAL_ADJUST)
- Sex (CL_SEX)
- Statistical operation (CL_STATISTICAL_OPERATION)
- Time format (CL_TIME_FORMAT)
- Time period – collection (CL_TIME_PER_COLLECT)
- Unit multiplier (CL_UNIT_MULT)

Istat

Relevant guidelines

Code IDs take values from uppercase A to Z, 0 to 9
and "_" only. No other characters should be used. Underscore ("_") is
generally used for the combination of codes (whether consecutive or not)

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.

Accented characters for a code are not allowed by the standard

Code names should be between 1 and 254 characters

The characters used should belong to the UTF-8 character set

it is recommended to use meaningful codes (in English) if possible,
pay attention not to use excessively long codes (impacts the maintainability of
identifiers/key series codes, size and costs of data files and databases)

Istat

Age classes

The basic units used for representing age are
"**Y**" (Year), "**M**" (Month), "**W**" (Week) and "**D**" (Day).

- The following set of standard operators is proposed:
- "**T**" for expressing ranges (e.g., from 4 to 9);
- "**_**" for the combination of two codes, whether consecutive or not;
- "**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".

⊕ from 15 years to 20 years excluding 16 years: **Y**15**T**20**X**16
⊕ over 30 years: **Y_GT**30 (here"_" is used to make a clear
distinction between the unit and the operator "GT")
⊕ less than 50 years: **Y_LT**50
⊕ two and three months: **M**2_3
⊕ four days or over: **D_GE**4
⊕ three years or less: **Y_LE**3

Istat

Generic codes

Code	Description	Annotation
_L	Local extension	to uniquely identify local extensions of SDMX code lists
_N	Non response	failure to obtain a measurement
_O	Other	residual information not contained in other categories of the code list
_S	Subtotal	used for expressing intermediate totals
_T	Total	used for expressing totals
_U	No data/unknown	failure to obtain a measurement
_X	Not allocated/ unspecified	the value for a particular variable falls outside the expected range
_Z	Not applicable	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.

Istat

Istat

AS-IS Structural Metadata

Structural metadata harmonization can be set as a long-term objective. Using the same terms according to a common language structure, from data collection to data dissemination, is a demanding task.

When it comes to structural metadata, the harmonization of the statistical information provided to final users is seen as top priority.

Therefore, following ISTAT experience,
the first clear and realistic steps can be taken
in the **dissemination** phase,
natural "bottleneck" for the different domains:
analyze the **Concepts** and the **Code lists**
used in the disseminated data so to understand
the issues to be tackled

Concept Scheme **+** Codelists

= DSD

Istat
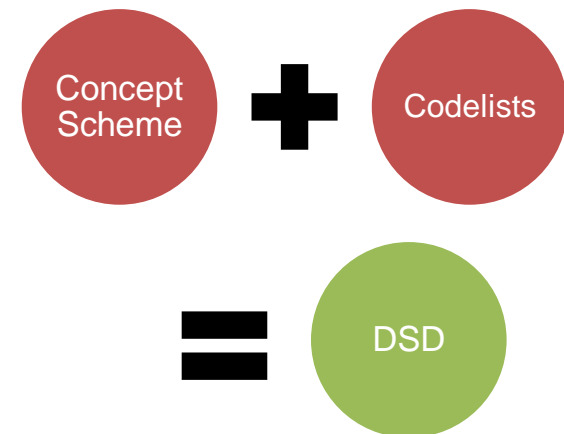
Work strands

1. Analysis and comparison of the different statistical processes using related variables

2. Discussion and standardization: variables, questions, guidelines (national and international context)

3. Proposal of a **unique classification**

4. Add "Special" categories + Aggregates needed (i) when gathering micro data and (ii) when disseminating macro data → **unique codelist**

Istat

(VN) Occupation

Gross domestic product at current prices by economic sector

CL_OCCUPATION (based on ISCO-08, International Standard Classification of Occupations, 2008 version)

(Version 1.0, formally adopted 14 March 2014)

	A	B	C
1	Code Value	Code Description	Notation 1 - Hierarchica
2	OC1	Managers	Hierarchical level 1
3	OC11	Chief executives, senior officials and legislators	Hierarchical level 2
4	OC111	Legislators and senior officials	Hierarchical level 3
5	OC1111	Legislators	Hierarchical level 4
6	OC1112	Senior government officials	Hierarchical level 4
7	OC1113	Traditional chiefs and heads of village	Hierarchical level 4
8	OC1114	Senior officials of special-interest organizations	Hierarchical level 4
9	OC112	Managing directors and chief executives	Hierarchical level 3
10	OC1120	Managing directors and chief executives	Hierarchical level 4
11	OC12	Administrative and commercial managers	Hierarchical level 2
12	OC121	Business services and administration	Hierarchical level 3
13	OC1211	Finance managers	Hierarchical level 4
14	OC1212	Human resource managers	Hierarchical level 4
15	OC1213	Policy and planning managers	Hierarchical level 4
16	OC1219	Business services and administration	Hierarchical level 4
17	OC122	Sales, marketing and development managers	Hierarchical level 3
18	OC1221	Sales and marketing managers	Hierarchical level 4
19	OC1222	Advertising and public relations managers	Hierarchical level 4
20	OC1223	Research and development managers	Hierarchical level 4
611	OC0	Armed forces occupations	Hierarchical level 1
612	OC01	Commissioned armed forces officers	Hierarchical level 2
613	OC011	Commissioned armed forces officers	Hierarchical level 3
614	OC0110	Commissioned armed forces officers	Hierarchical level 4
615	OC02	Non-commissioned armed forces officers	Hierarchical level 2
616	OC021	Non-commissioned armed forces officers	Hierarchical level 3
617	OC0210	Non-commissioned armed forces officers	Hierarchical level 4
618	OC03	Armed forces occupations, other ranks	Hierarchical level 2
619	OC031	Armed forces occupations, other ranks	Hierarchical level 3
620	OC0310	Armed forces occupations, other ranks	Hierarchical level 4
621			

Occupation *

Total 11 Selected 0

TOTAL
Leaders/managers
High level professionals
Mid-level professionals
Clerks
Personal services, protective workers and sales worker
Skilled agricultural, forestry and fishery workers
Craft and related trade workers
Plant and machine operators and assemblers
Unskilled occupations
Others

Search

☐ Beginning of row

Istat

(IT) Status in employment in the main job

II Level		I Level	
code	**description**	**code**	**description**
SAL	employee	SAL	employee
COLCO	co-ordinated log-term freelancer	SELF_NS	self-employed person without employees
POO	indipendent contractor		
LIBSD	professional without employees		
LIPSD	own account workers without employees		
IMP	entrepreneurs	SELF_S	self-employed person with employees
LIBCD	professional with employees		
LIPCD	own account workers with employees		
CFAM	family worker	CFAM	family worker

Level 2 is more disaggregated. This is the level to be considered as the standard of acquisition for surveys that need to detect in detail the characteristics of the work. Level 1 is more aggregated. This is the minimum level of acquisition: it is to be used only in very rare cases, only by surveys that do not need to collect very detailed information about work.

Level 2 links up with Level 1: Level 2 detects employees and family workers but distinguishes the independents by unpacking them among those who have a coordinated and continuous collaboration (with or without a project), occasional workers, entrepreneurs, freelancers (with and without employees) , self-employed workers (with and without employees). This classification must be used by those surveys that have to investigate the characteristics of the work in detail.

Istat

(IT) Educational level

Acquisition level 1		Acquisition level 2		Acquisition level 3	
Code	Category	Code	Category	Code	Category
NED	no educational degree	NED	no educational degree	NED	no educational degree
PSE	final assessment (primary school)	PSE	final assessment (primary school)	PSE	final assessment (primary school)
LSE	diploma of lower secondary education	LSE	diploma of lower secondary education	LSE	diploma of lower secondary education
USV	diploma of upper secondary education (2-3 years)	USV_IFP	diploma of upper secondary education (2-3 years) and Vocational training qualification (three-year courses)/ Professional diploma (fourth year)	USE_IF	diploma of upper secondary education and Certification of higher technical specialisation
IFP	vocational training qualification (three-year courses)/ Professional diploma (fourth year)				
USG	diploma of upper secondary education (4-5 years)	USG_IFTS	diploma of upper secondary education (4-5 years) and Certification of higher technical specialisation		
IFTS	Certification of higher technical specialisation				
ITS	Diploma of Higher Technical	BL	bachelor's degree or first level academic diploma; Diploma of Higher Technical; University Diploma	BMRD	bachelor's degree or first level academic diploma; Diploma of Higher Technical; University Diploma; Master's degree or second level academic diploma (and Academy Conservatory old programme); Research Doctorate (PhD)/ Advanced research academic diploma
CDU	University diploma				
L	Bachelor's degree				
FLA	Fine arts, Drama, Dance and Music First level academic diploma (Bachelor's degree)				
LS	master's degree	ML	Master's degree or second level academic diploma (and Academy Conservatory old programme)		
LD	master's degree				
SLA	Fine arts, Drama, Dance and Music Second level academic diploma (Master's degree)				
ADO	academy Diploma old programme				
RDD	research Doctorate (PhD)/ Advanced research academic diploma	RDD	research Doctorate (PhD)/ Advanced research academic diploma		

Isced-8   Isced-7   Isced-6   Isced-5   Isced-8   Isced-7   Isced-5,6   Isced-5,6,7,8

CODE LIST		
Code	Category	Used by
IL	Illiterate	Pop Census
LBNA	Literate but no formal educational attainment	Pop Census
NED	no educational degree	R&D
PSE	final assessment (Primary school)	Absolute poverty
NP	no educational degree, final assessment (Primary school)	EU-SILC
LSE	diploma of lower secondary education	

Istat

In dissemination…

Imagine two or more statistical processes in your NSI taking different alternatives for, say, age groups.

In the ideal world, the long-term integrated optimizing solution is the use of a unique Code list named "Age / Duration".

Managers of every domain would select their subset of items from that unique Code list without creating anything domain-specific on their own.
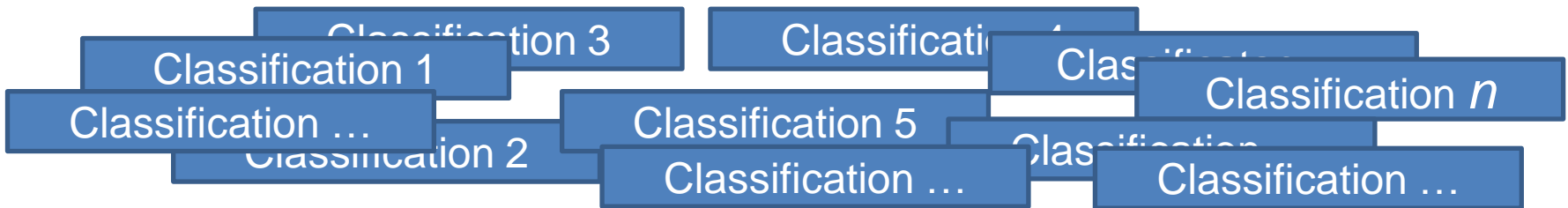
Advantages:
i.   internally, integration among statistical domains
ii. internally, no proliferation of objects
iii. externally, common standards → streamlined metadata exchange between organizations.

Istat

In the meantime, in the real world…

Survey 1 [Inactivity rate]	Survey 2 [Enterprise]	Survey 3 [Unmarried person]
In the Questionnaire		
Age: $(\mathbf{0}, \infty)$	**Age:** $(\mathbf{0}, \infty)$	**Age:** $(\mathbf{0}, \infty)$
On the Website		
01: 15-24 years **02**: 25-34 years **03**: 35-49 years **04**: 50-64 years **05**: 15-64 years	**17.1**: up to 24 months **17.2**: from 3 to 5 years **17.3**: from 5 to 10 years **17.4:** from 10 to 15 years **17.5**: 15 years and over **17.6**: all items	**A**: less than 35 years **B**: 35_49 years **C**: 50_59 years **D**: 60 years and more **E**: total

From the single need to…

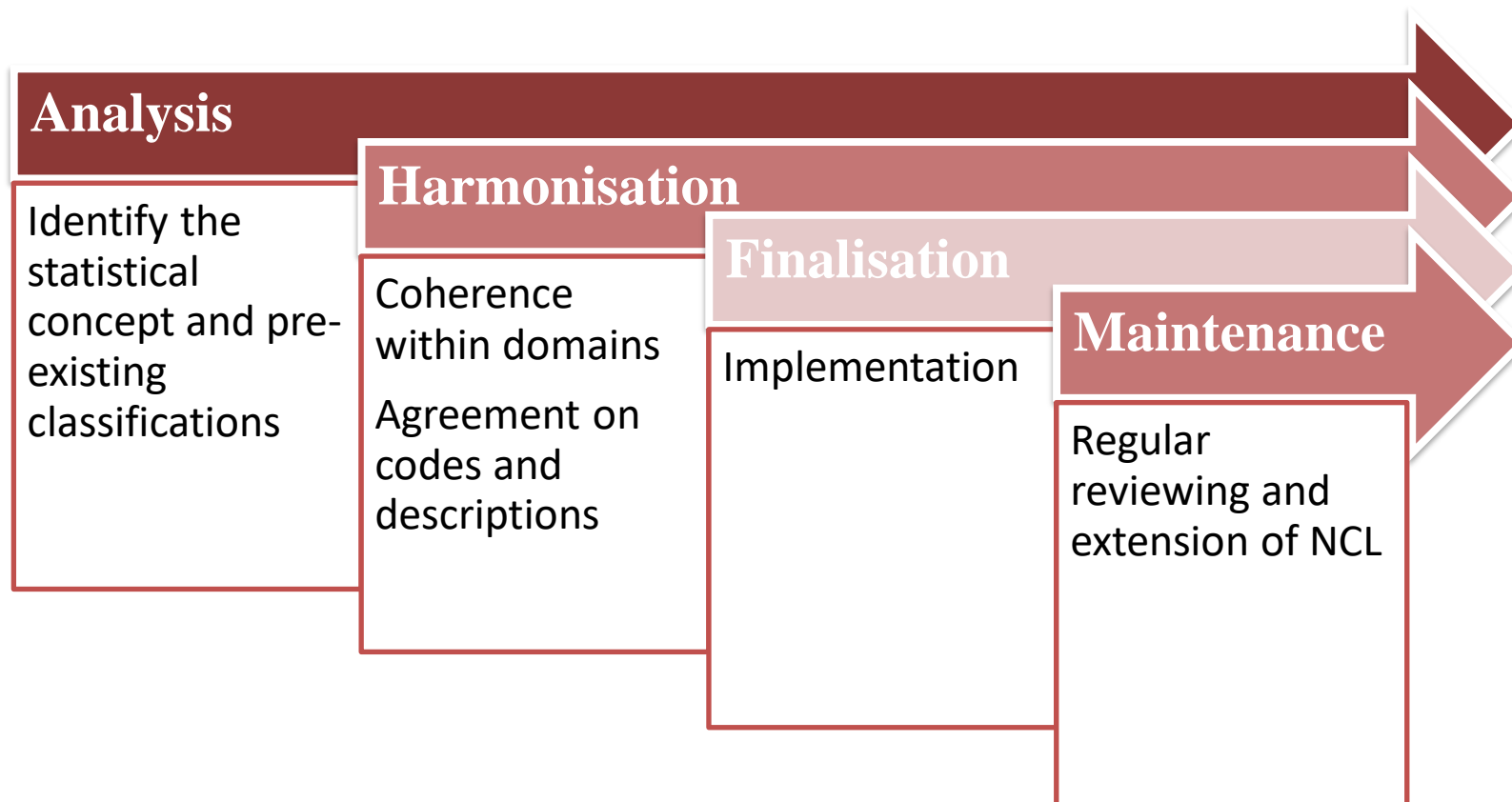Identify the statistical concept "AGE" throughout your *Statistical Programs*

Classification 3
Classification 1
Classification …
Classification 2
Classification 5
Classification …
Classification n
Classification …

Objective:

- ✓ coherence within statistical domains

- ✓ agreement on codes and labels

**SCL
Age/Duration**

A new Standard Code List (SCL) is born!!!

Istat

The process of standardisation

From multiple, disconnected classifications to a unique NCL

Analysis

Identify the statistical concept and pre-existing classifications

Harmonisation

Coherence within domains

Agreement on codes and descriptions

Finalisation

Implementation

Maintenance

Regular reviewing and extension of NCL

Istat

Single Corporate Dissemination DB

By now,

the NCL "AGE/DURATION" has been recalled 183 times, in 148 dataset.

It has 356 items

Code	Text
D0	0 days
D0-6	0-6 days
M10	10 months
M11	11 months
NAP	not applicable
NDIV	not divisible
NSP	no response
TOTAL	total
UNK	age unknown
Y_GE14	14 years and over
Y_GE15	15 years and over
Y_UN16	until 16 years
Y_UN17	until 17 years
Y0-13	0-13 years
Y0-14	0-14 years
Y35-39	35-39 years
Y35-44	35-44 years
Y35-49	35-49 years
Y35-54	35-54 years

Advantages

less workload on production units: less mapping, less conversions,

more comparability across domains

the adoption of the harmonised codes into the data production chain would reduce the need for transcodification and the risk of errors.

statistical outputs harmonised, comparable, consistent

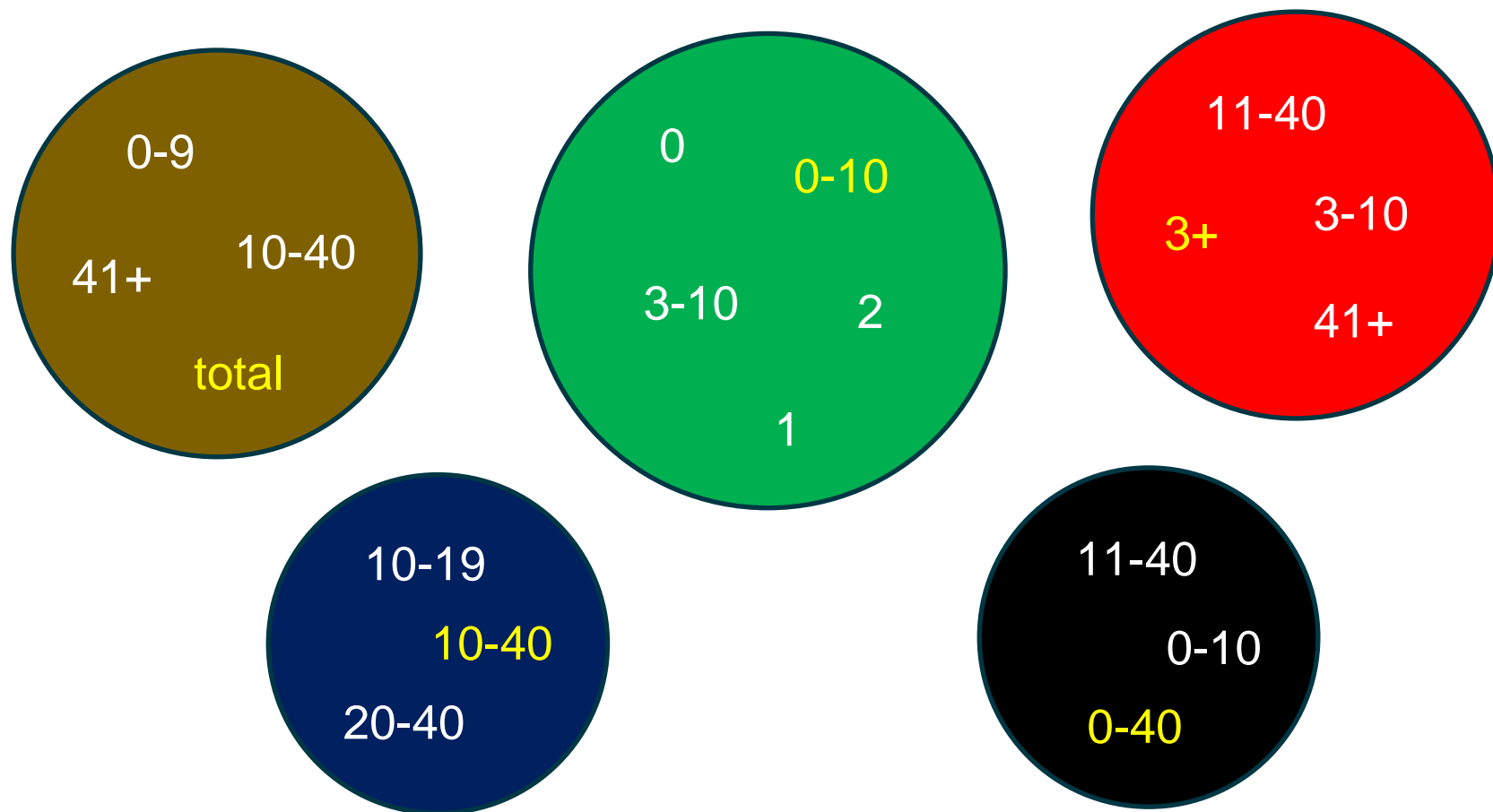**a choice over NCL has a great impact on the efficiency of data sharing**

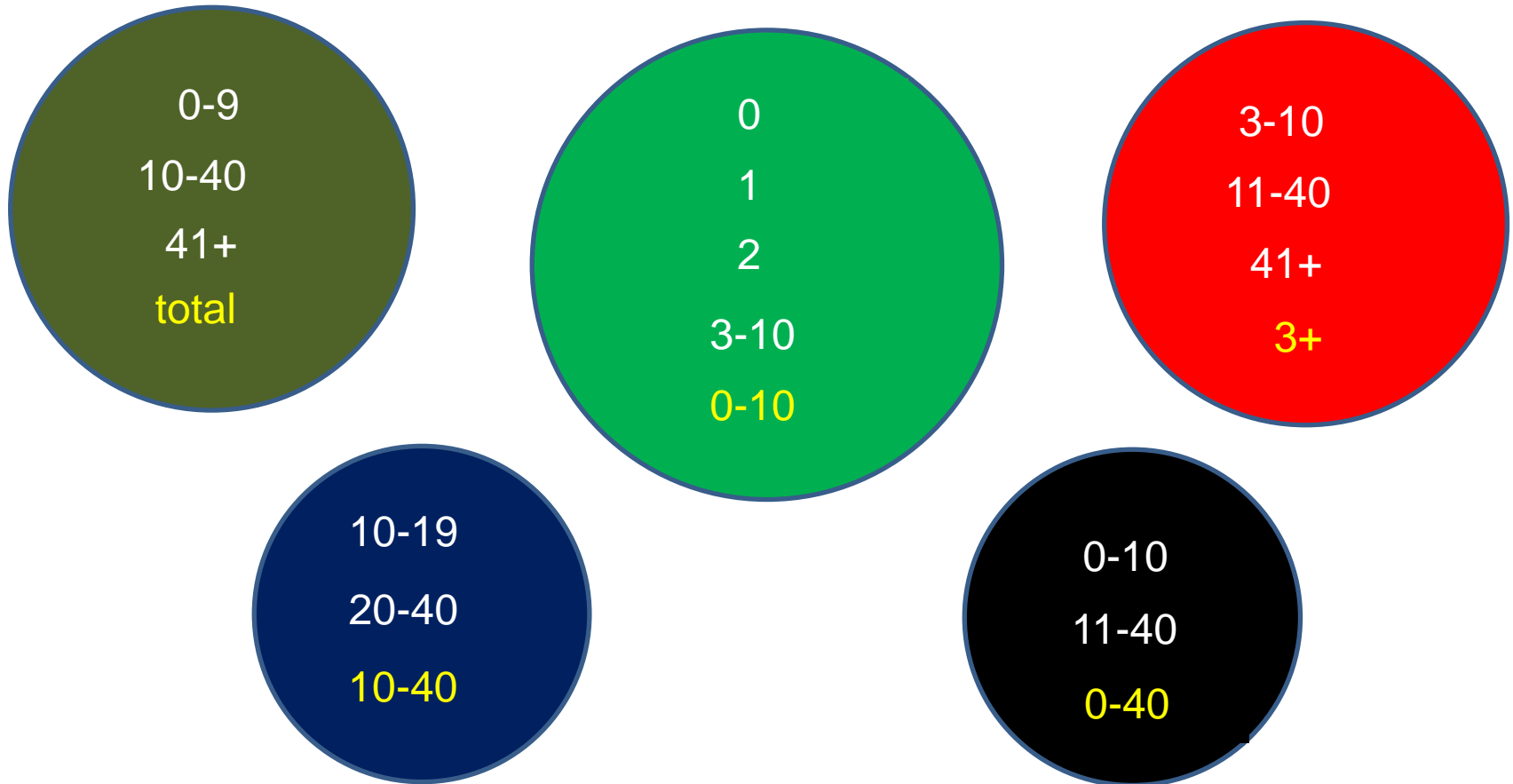Exercise - How to order a unique Code list

Istat

Starting point: five ~~classifications~~ code lists

These are the Age *items* in the different *Statistical Programs*

0-9

10-40

41+

total

0

0-10

3-10

2

1

11-40

3+

3-10

41+

10-19

10-40

20-40

11-40

0-10

0-40

Istat

Easy ordering

It's easy to sort the items, if you keep them separated…

0-9
10-40
41+
total

0
1
2
3-10
0-10

3-10
11-40
41+
3+

10-19
20-40
10-40

0-10
11-40
0-40

Istat

Amman, we have a problem
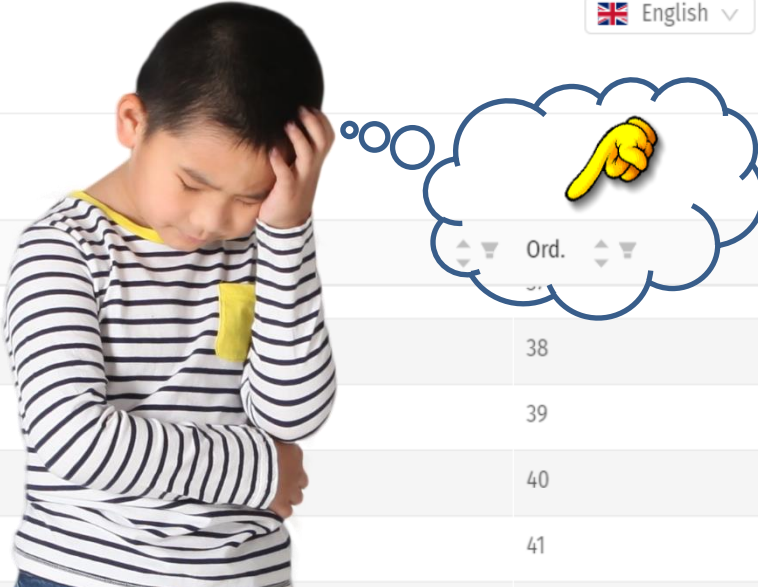
Obviously, things get complicated about the ORDER,
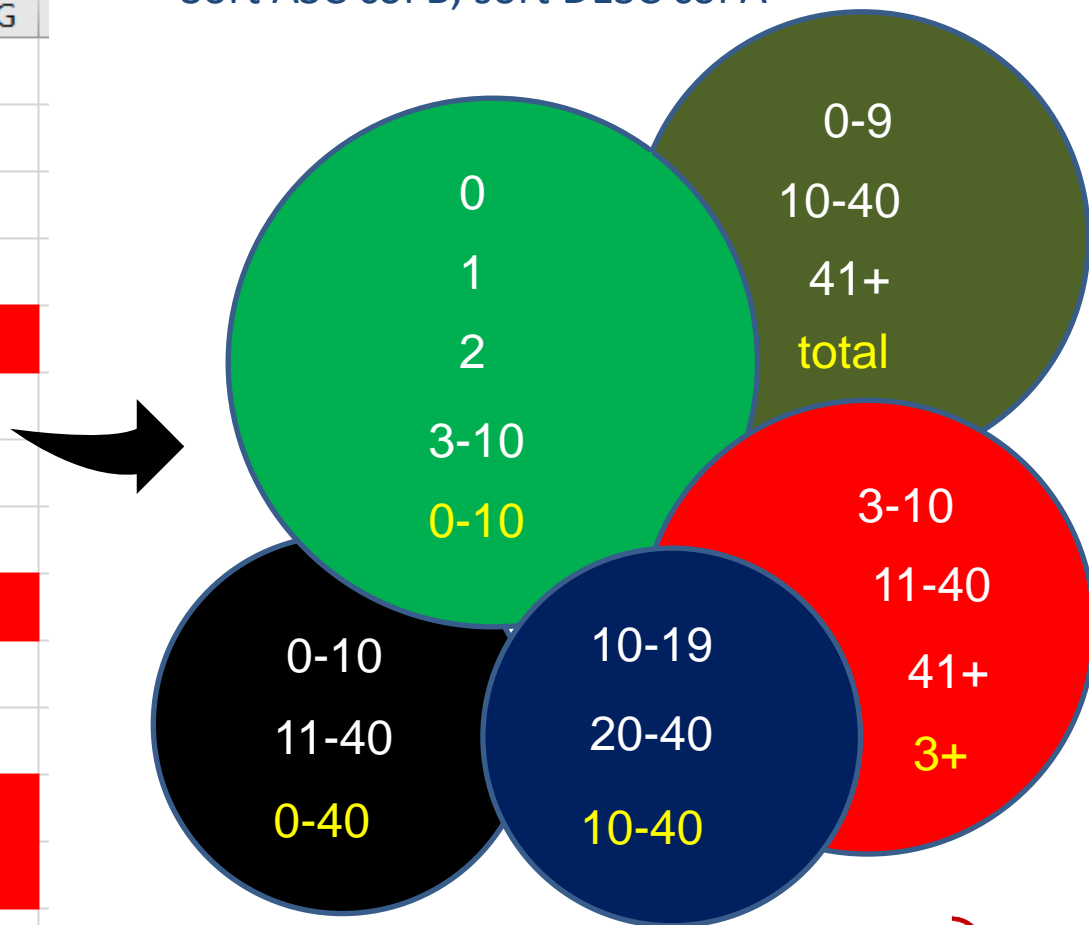when you collect hundreds of Age *items* in the same flat *code list*...

Question

Is it possible to sort CL_AGE *items* in such a way that every possible subset extracted is automatically sorted? **YES**

CL_AGE

Istat
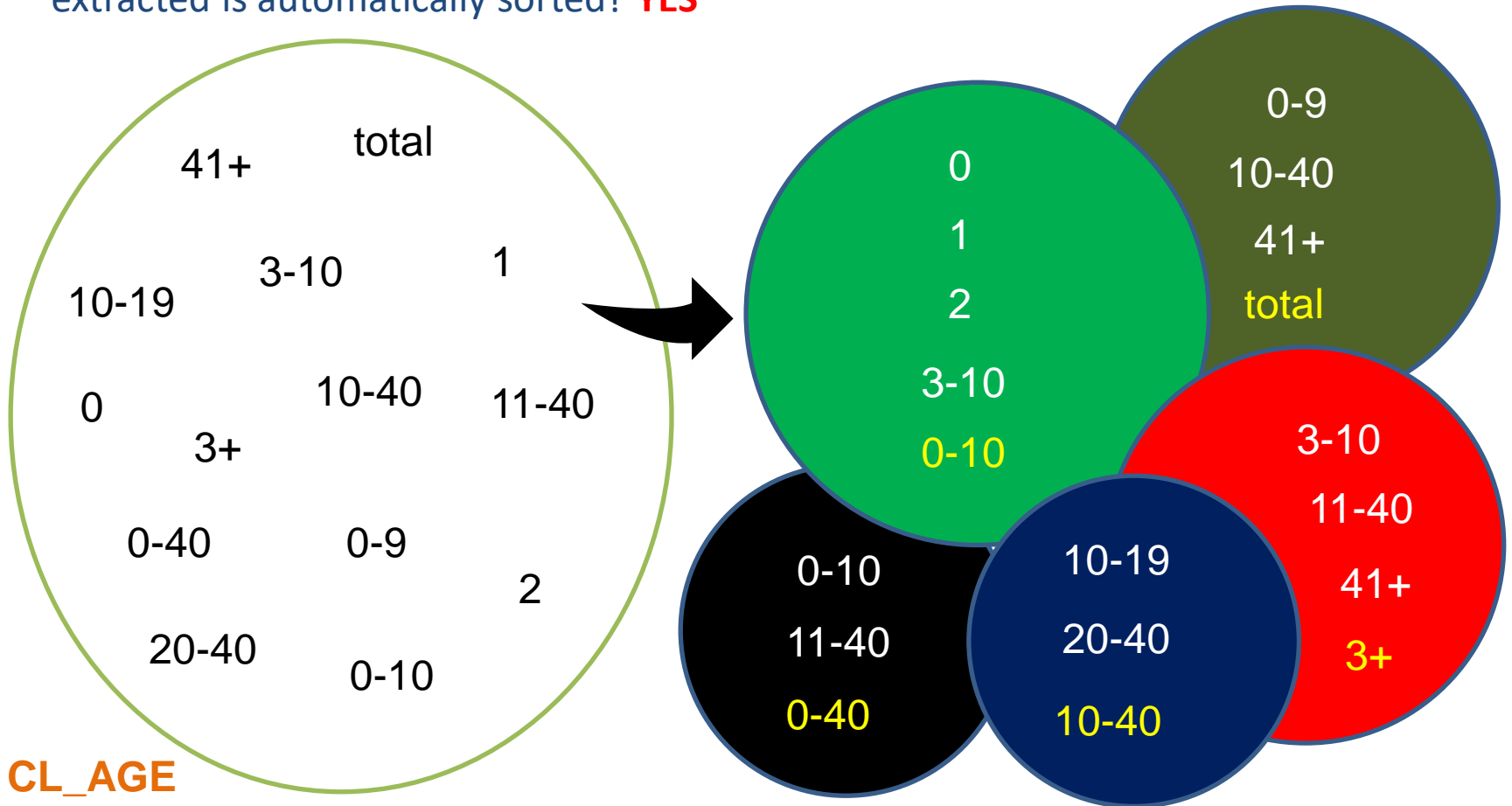
Simple as that...

- At bottom, sort DESC open-ended intervals
- Sort ASC col B, sort DESC col A

Andrea Bruni
anbruni@istat.it