“Strengthening the capacity of Jordan’s Department of Statistics”

Activity 1.3.6: Standardized production process and the role of metadata

Structural metadata governance at Istat: harmonisation table

Mr. Andrea Bruni

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

the need for harmonization of metadata

Today, integrating different sources means understanding the metadata associated with different data sets, defined autonomously in different *Statistical Programs*, often using different terms with the same meaning or same terms with different meaning.

So, why standardize metadata?

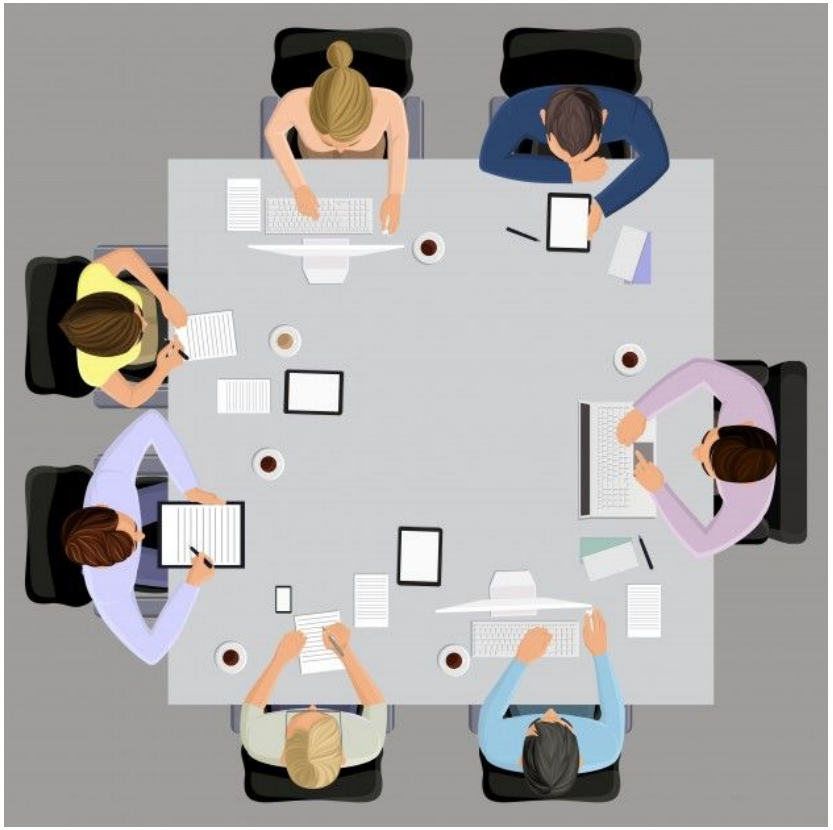
- to make the data produced in the different phases homogeneous in terms and definitions, in order to broaden the amount of data that can be integrated and analyzed jointly.
- to identify data more easily by appropriate search functions. The aspect related to the search functionality is not negligible, neither for the employees of the Institute, nor among external users. External users formulate a question of information which is always linked to a metadata.

The **Institute's standard metadata** are shared, reusable in in different phases and in different Statistical Programs

Metadata table

Table dedicated to the harmonization of metadata

The objective of the table is to define the **Institute's standard metadata** to be reused in every Statistical Program, from data collection to the dissemination phase.



Step by step

Phases necessary to define the Institute's standard metadata

collection of useful materials, drafting of a standard metadata proposal (including definition, coding, etc.) carried out by thematic experts

the standard metadata proposal is analysed by the Metadata Table about its applicability in the different stages of the production process (data acquisition, data collection, dissemination);

the "validator" approves the standard metadata proposal that assumes the role of the **Institute's standard metadata**

Actors and roles

keyword: involve your “partners”

Step	In charge	Consulting
INVESTIGATION	METADATA REFERENT on behalf of the thematic experts pool	Production units
DISCUSSION	METADATA EXPERT on behalf of the *Metadata Table*	Thematic experts (and their metadata referent)
VALIDATION	DIRECTOR responsible for the metadata	Other directors (and metadata expert)

1st phase: investigation

The **Metadata Referent** collects all the material necessary to define a standard metadata. He/is a thematic expert on metadata, generally engaged in data production activities. Knows the different metadata application strategies in Statistical Programs and in the different stages of production.

For metadata more relevant, for example for official classifications, is the NSI representative who participates in the tables for the definition of metadata in national or international contexts (Eurostat, UN, etc).

He/She listens to the production units (and cross-domain specialists) and ensures that all the needs of the production processes, all phases of the cycle are being considered, but remains the unique reference person of the *Metadata Table*.

1st phase: investigation

He/She prepares a preliminary standard metadata proposal, starting from the recommendations defined at international level, from the regulations, from the NSI standards on metadata, drawing up the appropriate **template**.

The product of the preliminary phase is the standard metadata proposal accompanied by a **technical report**.

The investigation phase ends by sending the documents to the metadata table.

Technical report

<table>
<tr><td>Concept</td><td>Highest level of education attained</td></tr>
<tr><td>Identifiable Attribute_ID</td><td>HATLEVEL</td></tr>
<tr><td>Identifiable Attribute_Version</td><td>dd/mm/yyyy</td></tr>
<tr><td>Metadata Referent</td><td>Name Surname</td></tr>
<tr><td>Thematic Experts consulted</td><td>Person 1, Person 2, …, Person n</td></tr>
<tr><td>Unit type</td><td>Individuals</td></tr>
<tr><td>Constraints</td><td>The question is asked to different groups, regardind age:
EU- LFS: 15+ years, POPCENSUS 9+ years, HBS 9+ years
AES: 18-74 years, EU-SILC 16+, AVQ 6+

EU level - transmission of microdata
<table>
<tr><td rowspan="2">Micro-data collection</td><td colspan="3">Level of detail</td></tr>
<tr><td>High[1]</td><td>Medium[2]</td><td>Low[3]</td></tr>
<tr><td>EU-SILC</td><td>X (age 16-34)</td><td>X (age 35+)</td><td></td></tr>
<tr><td>EU-LFS</td><td>X (15+)</td><td></td><td></td></tr>
<tr><td>HBS</td><td></td><td></td><td>X (15+)</td></tr>
<tr><td>AES</td><td>X</td><td></td><td></td></tr>
<tr><td>EHIS</td><td></td><td></td><td>X</td></tr>
<tr><td>HETUS</td><td></td><td></td><td>X</td></tr>
<tr><td>ICT HH</td><td></td><td></td><td>X</td></tr>
</table>
[1] 3 digit ISCED
[2] 2 digit ISCED
[3] 1 digit ISCED

The indication is to make the populations homogeneous where possible in all surveys. As-is situation: the harmonized population is the intersection of the populations investigated in the individual surveys - Population 18-74 years.</td></tr>
<tr><td>International reference</td><td>International Standard Classification of Education (ISCED), 2011 version</td></tr>
<tr><td>Levels</td><td>3</td></tr>
<tr><td>Levels in data collection</td><td>1</td></tr>
</table>

Technical report

Remarks	In data collection: recommended to use the *Classification items* listed in the *Statistical Classification*. In *Statistical Programs* not having direct interest to the *Concept* in their to the cognitive objectives, a more aggregate level of acquisition is expected. The Population Census uses ISCED plus a focus dedicated to (...). EU-LFS, EU-SILC, AES, Aspects of Daily Life, HBS use ISCED plus investigation of the presence of university masters.HBS not using: IFP, IPS. In data collection, the strategy of investigating the highest qualification achieved through a sequence of two or more *Questions* is often adopted. A first *Question* identifies the qualification in a less fine way (...). Afterwards, detailed *Questions* are inserted for (...) respectively. In EU-LFS, EU-SILC, AES *Questionnaires*, post-graduate qualifications also include the different types of masters (which do not actually constitute a higher qualification than a degree). The need to standardize the *Designation* of the *Code* assigned to each *Classification Item* by the various *Statistical Programs* emerges.
Recommended *Instance Question*	What is the highest educational qualification that [...] has achieved? Are qualifications awarded by vocational education and training courses (IFP, IFTS, ITS) included? However, the question must be interpreted, according to the instructions of the questionnaire, as the highest qualification up to the qualification that allowed access to post-graduate studies. Possession of postgraduate qualifications is investigated in a subsequent question: Did you hold a postgraduate, postgraduate AFAM academic degree or PhD?
Guidelines	Persons with two or more qualifications of the same level must indicate: • the most recent (Eurostat, EU-SILC, EU-LFS). • the one deemed most important in relation to any professional activity exercised (CENSUS OF THE POPULATION). Those who are attending a course of study must not indicate the diploma they will subsequently obtain, but the one they already have (Eurostat, EU-SILC, EU-LFS). People (in particular foreign citizens) who have attained the highest level abroad must select the *Classification Item* relating to the corresponding qualification in Italy (even if not legally recognized) (CENSUS OF THE POPULATION, EU-SILC, EU-LFS, Aspects of Daily Life, HBS) (and more)
Documentation taken into consideration	ISCED 2011, LC-LEGAL 43-17 Standardised variable, Handbooks and *Questionnaires* of the main *Statistical Programs* regarding households/individuals

Template

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

2nd phase: Discussion and Delivery

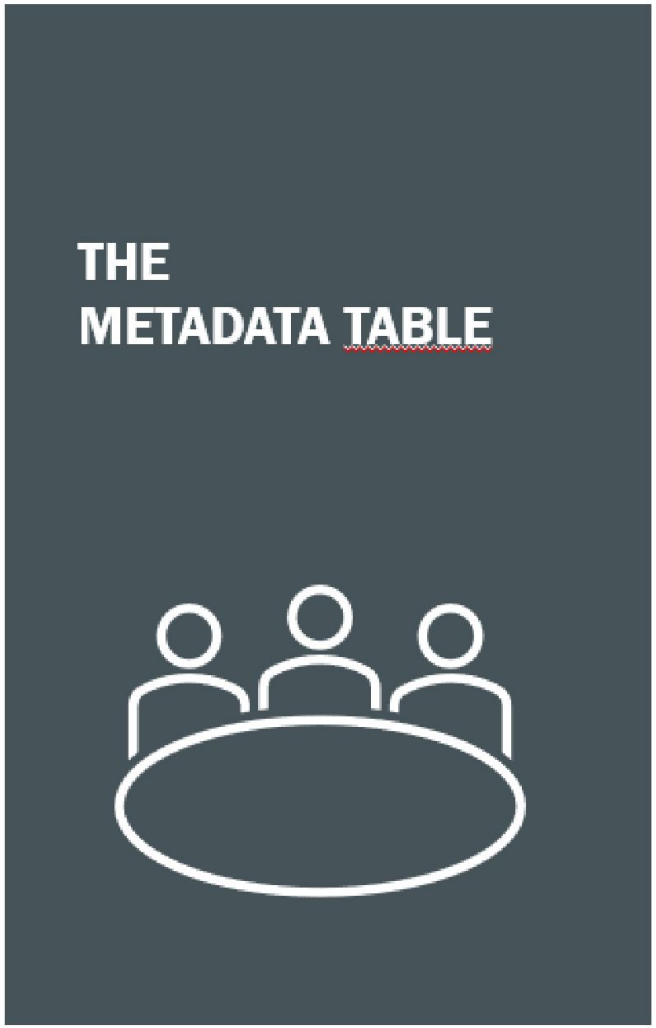
The metadata proposal is discussed together with the **Metadata Referent** in plenary session by the Metadata Table, which takes into account the metadata management along the phases of the production processes, in order to verify both (i) compliance with the standards and (ii) the real possibility of use.

The members of the table are the metadata experts + colleagues who deal with the harmonization of metadata in different production units.
From time to time, experts in the classifications subject to harmonization are invited to collaborate in defining the standards.

In this phase, critical issues other than those indicated in the preliminary phase may be highlighted.

Proposals for tests/experiments may also be made for evaluating the impact of the implementation.

Finding agreements



It is where the metadata is discussed and sent for validation.
The documents produced in the preliminary phase are checked in the light of the standards defined and the overall needs of the data life cycle.

The **technical report** is discussed and, considering the needs related to the use of metadata in the production phases, issues of concern are highlighted and addressed

In the event of no agreement being reached, the investigation is sent to **Directors** pointing out the critical problems encountered.

Use of the standard metadata

Once the standard metadata is defined, its use to describe the data, in all phases, must be implemented.

Task of the Metadata Table: **promote of the use** of standard metadata from the questionnaires to the data present in the various archives of the Institute, to data tables disseminated.

Task of the Metadata Table: **support sectors** by finding common strategies for the adaptation or evolution of metadata.

The metadata harmonization activity does not end, therefore, with the definition of the standard

Resources and time are needed to implement the standard metadata.
Versioning: every metadata must be equipped with its validity period.
Who reports any changes, updates or additions to the Metadata Harmonization Table? And how?

Commitment

Deviation from the standard must be adequately motivated.

The Metadata System will produce reports on the degree of use of standard metadata.

These reports will be sent to **competent directors** in order to be able to define appropriate strategies that favor the use of metadata standard.

Workshops, capacity building, opportunities, communication on the intranet (transparency)

ACTIVE role of metadata

The Metadata System (MS) must act as a “once for all purposes“ repository, exposing existing metadata. Then, each data management tool must acquire metadata from MS.

MS must provide appropriate and effective **services** with the aim of facilitating the re-use of metadata in other processes and at all GSBPM stages (e.g., data acquisition, validated data storage, data dissemination)

If the MS is not provided with the metadata felt necessary, a procedure for requesting a new metadata must be activated.

Work in progress…

METAstat: the new Istat Metadata System

The importance of interoperability, which can be pursued primarily through the use of statistical models and standards existing at an international level (GSBPM, GSIM, etc.), has clearly emerged in the context of the National Data Catalogue and can be achieved by having two fundamental infrastructures available: a complete and transversal metadata system together with ontologies and controlled vocabularies.

Istat is currently working on the creation of METAstat, the new institutional system for the documentation of metadata, processes and statistical products.

METAstat is designed not to be a passive catalogue of metadata, to be fed ex post, but must have an active role in providing production services with the concepts (represented by metadata) on which to structure the data to be produced (metadata driven). It will enter the production processes already in the design phase of the survey and will have to be integrated into the production processes. METAstat goal is not to be a catalogue for purely documentary purposes, but to provide active support to simplify and automate production processes, as well as to increase the reliability, consistency and timeliness of the data produced (quality principles). In this way the sharing (internal or external) of the data produced will be simplified and facilitated, because these data are already structured on shared and certified metadata from birth.

It is clear how crucial the aspects of governance and shared rules are before the development of the system.

Governance principles

Key CHALLENGE:
ensure that metadata are captured as early as possible, and stored and transferred from phase to phase alongside the data they refer to

Keep TRACK of the versioning, the processes defining metadata,
the processes just reusing metadata

MAINTENANCE:
ensure non-obsolescence, technological and semantic

HARMONIZATION:
never forget international standards (GSBPM, **GSIM**, **SDMX**)

Andrea Bruni
anbruni@istat.it