

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

*Activity 1.3.6: Standardized production process and the role of metadata*

# Metadata overview

**Mr. Andrea Bruni**

ISTAT | DIRECTORATE FOR EXTERNAL RELATIONS, INTERNATIONAL AFFAIRS, PRESS OFFICE  
AND NATIONAL STATISTICAL SYSTEM COORDINATION



Delegation of the European  
Union to Jordan



# Definition of Metadata

Greek **μετά** «AFTER, BEYOND» + latin **datum** «INFORMATION»

ISO/IEC 11179 on Metadata Registries

Data that **defines** and **describes** other data.

## EUROSTAT

Information that is needed to be able to **use** and **interpret** statistics.

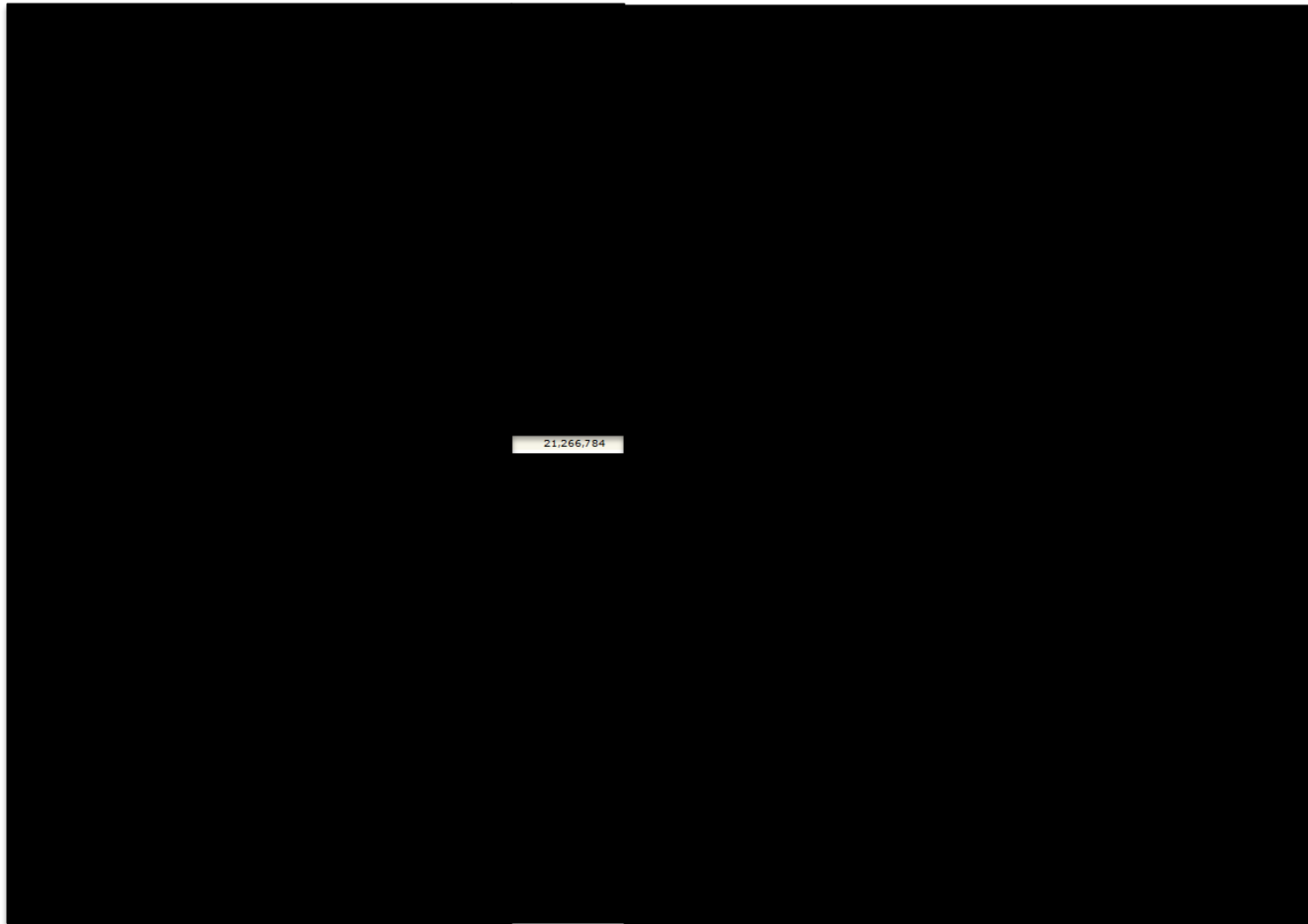
Metadata **describe** data by giving definitions of populations, objects, variables, the methodology and quality.

Cambridge dictionary

Information that is given to **describe** or **help you use** other information

A distinction is generally made between **structural** and **reference** metadata

## Identify metadata



# A quality report

[http://ec.europa.eu/eurostat/cache/metadata/en/tour\\_occ\\_esms.htm](http://ec.europa.eu/eurostat/cache/metadata/en/tour_occ_esms.htm)



## Occupancy of tourist accommodation establishments (tour\_occ)

Reference Metadata in Euro SDMX Metadata Structure (ESMS)  
Compiling agency: Eurostat, the statistical office of the European Union

Eurostat metadata	
<b>Reference metadata</b>	
1. <a href="#">Contact</a>	
2. <a href="#">Metadata update</a>	
3. <a href="#">Statistical presentation</a>	
4. <a href="#">Unit of measure</a>	
5. <a href="#">Reference Period</a>	
6. <a href="#">Institutional Mandate</a>	
7. <a href="#">Confidentiality</a>	
8. <a href="#">Release policy</a>	
9. <a href="#">Frequency of dissemination</a>	
10. <a href="#">Accessibility and clarity</a>	
11. <a href="#">Quality management</a>	
12. <a href="#">Relevance</a>	
13. <a href="#">Accuracy</a>	
14. <a href="#">Timeliness and punctuality</a>	
15. <a href="#">Coherence and comparability</a>	
16. <a href="#">Cost and Burden</a>	
17. <a href="#">Data revision</a>	
18. <a href="#">Statistical processing</a>	
19. <a href="#">Comment</a>	
<a href="#">Related Metadata</a>	
<a href="#">Annexes (including footnotes)</a>	

National metadata			
<b>National reference metadata</b>			
National metadata produced by countries and released by Eurostat			
<a href="#">Belgium</a>	<a href="#">Bulgaria</a>	<a href="#">Czech Republic</a>	<a href="#">Denmark</a>
<a href="#">Germany</a>	<a href="#">Estonia</a>	<a href="#">Ireland</a>	<a href="#">Spain</a>
<a href="#">France</a>	<a href="#">Croatia</a>	<a href="#">Italy</a>	<a href="#">Latvia</a>
<a href="#">Lithuania</a>	<a href="#">Luxembourg</a>	<a href="#">Hungary</a>	<a href="#">Malta</a>
<a href="#">Netherlands</a>	<a href="#">Austria</a>	<a href="#">Poland</a>	<a href="#">Portugal</a>
<a href="#">Romania</a>	<a href="#">Slovenia</a>	<a href="#">Slovakia</a>	<a href="#">Finland</a>
<a href="#">Liechtenstein</a>	<a href="#">Switzerland</a>	<a href="#">Montenegro</a>	<a href="#">Serbia</a>

For any question on data and metadata, please contact: [EUROPEAN STATISTICAL DATA SUPPORT](#)

[Download](#)

1. Contact <span style="float: right;"><a href="#">Top</a></span>	
1.1. Contact organisation	Eurostat, the statistical office of the European Union
1.2. Contact organisation unit	G3: Short-term Business Statistics and Tourism
1.5. Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update <span style="float: right;"><a href="#">Top</a></span>	
2.1. Metadata last certified	08/06/2017
2.2. Metadata last posted	08/06/2017
2.3. Metadata last update	08/06/2017

3. Statistical presentation <span style="float: right;"><a href="#">Top</a></span>	
3.1. Data description	
Accommodation statistics are a key part of the system of tourism statistics in the EU and have a long history of data collection. Annex I of the <a href="#">Regulation (EU) 692/2011 of the European Parliament and of the Council</a> deals with accommodation statistics and includes 4 sections focusing on accommodation statistics of which sections 1 and 2 include the requirements concerning rented accommodation (capacity and occupancy respectively). Data are collected by the competent national authorities of the Member States and are compiled according to a harmonised methodology established by EU regulations before transmission to Eurostat. Most of the time, data are collected via sample or census surveys. However, in a few cases data are compiled from a demand-side perspective (i.e. via visitor surveys or border surveys). Surveys on the occupancy of accommodation establishments are	

## Structural metadata

used to identify, define and accompany statistical data

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

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.

To illustrate: The number 3,651,881 is meaningless unless one is provided with its accompanying structural metadata, namely that this is the total number of women in Bulgaria on the 1st January 2017.

## Reference metadata

More general nature, to assist with the interpretation of the data:

- include explanatory texts on the context of the statistical data
- describe the statistical concepts and methodologies used for the collection and generation of data
- provide information on output quality

They can be decoupled from the data; this means, they can be generated, collected, stored, exchanged and disseminated separately from the statistics to which they refer, without being embedded in the data message.

In other words, those metadata are normally linked to the object by a simple “reference” to the object.

Very often, these metadata are associated not with specific observations or series of data, but with entire collections of data or even with the institutions providing the data.

## Reference metadata

Preferably, reference metadata should include all the following:

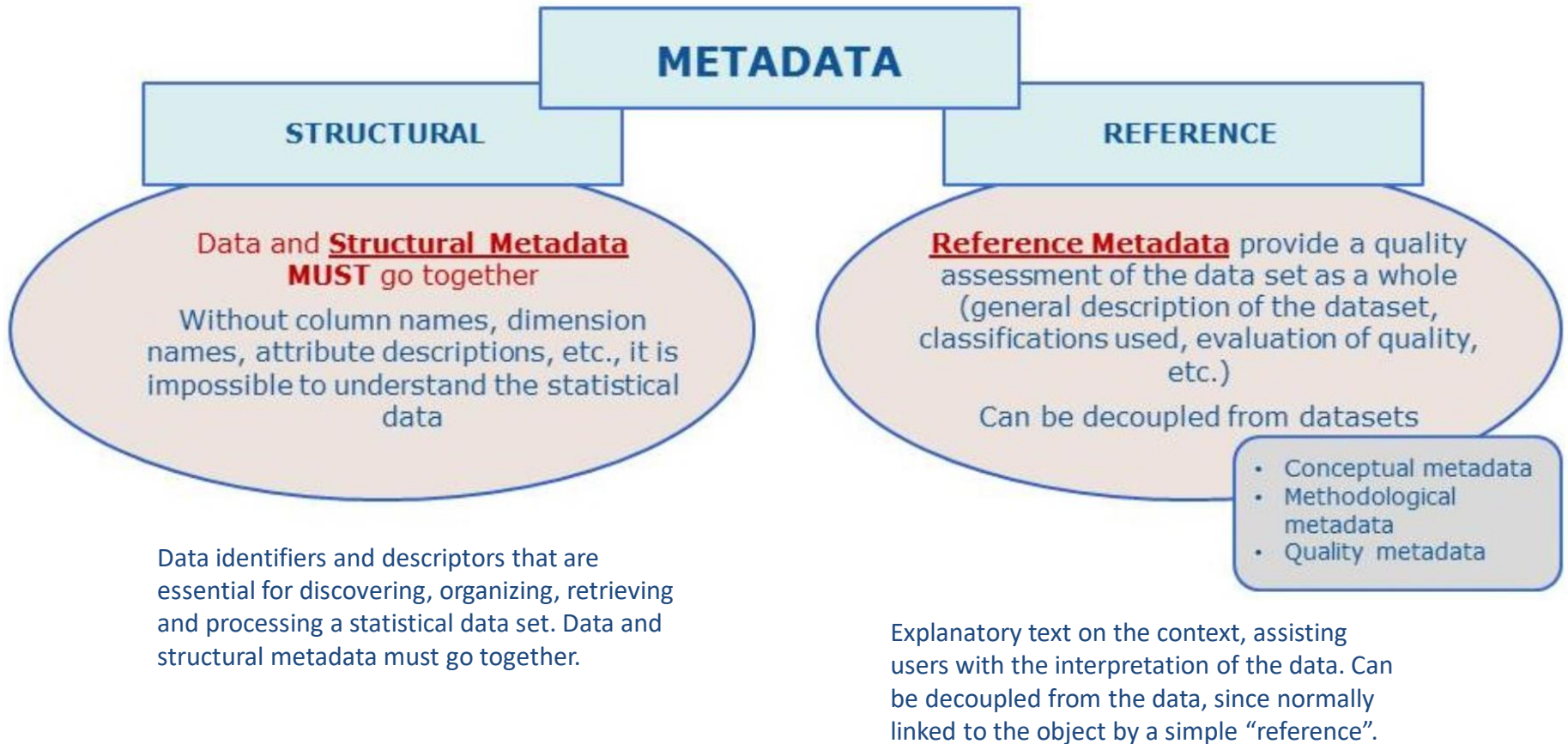
- a) **conceptual** metadata, describing the concepts used and their practical implementation, allowing users to understand what the statistics are measuring and, thus, their fitness for use;
- b) **methodological** metadata, describing methods used for the generation of the data (e.g., sampling, collection methods, editing);
- c) **quality** metadata, describing the different quality dimensions of the resulting statistics (e.g., timeliness, accuracy).

Information on **concepts** are connected to relevance, comparability and coherence of statistics, that are quality dimensions.

Information on **methodologies** applied during the statistical process are related to process quality.

Information on output **quality** equals quality dimensions by definition.

# Metadata diagram





## The need for metadata

The provision of metadata with statistics outlining concepts, definitions and describing methods used in collection, compilation, transformation, revision practices and dissemination of statistics, etc., is an essential function of all statistical agencies.

Lend methodological transparency (UN Fundamental Principles of Official Statistics) to economic, social and population statistics so that the typical end-user can make an informed assessment of their usefulness and relevance to his or her purpose.

The provision of metadata is therefore an unavoidable responsibility of all statistical agencies and one that requires adequate planning and resources.

## In practice

DATA AND METADATA REPORTING AND PRESENTATION HANDBOOK - ISBN 92-64-03032-8 - © OECD 2007

All statistical agencies should:

- compile metadata required for users to understand the strengths and limitations of the statistics it describes
- establish active linkage of metadata to the statistical tables and graphs they describe and vice versa
- by using a common set of terminology, structure a layered presentation of methodological items (or metadata elements), progressing from summary metadata to more detailed metadata
- provide contact persons or email addresses where further information may be obtained. In some organizations the “contact” would be a generic corporate contact point or referral service for all client enquiries
- keep their metadata up-to-date, incorporating the latest changes in definitions, classifications and methodology, etc.
- provide metadata not only in the national language(s) but also, where resources permit, in a common language such as English

## Issues of concern

There are two broad sets of issues with respect to the metadata management, namely the

**Accessibility of the metadata:** actual availability of metadata on NSI website, organisation on the web, provision of search facilities (local engine based on free text search), linkage to data.

**Content of the metadata:** differences in the actual statistical methodological elements described in metadata within/cross statistical domains. In some instances, the problem is merely one of terminology where the same term can have different meanings or different terms can have the same meaning. In other cases, the actual metadata is different. Any meaningful methodological comparisons is made a time consuming and costly exercise.

## Metadata governance

Establish a centrally managed and controlled statistical metadata framework and process that drives standard definitions, concepts and classifications to ensure data accuracy and common data interpretation.

The standardised catalogue of metadata (centrally managed and controlled) must be shared with all statistical staff so that they can conform to a set of predefined metadata definitions.

When applied broadly, metadata can also help in IT-based solutions for automating standardized survey plans, survey methods and survey questionnaires.

Make it a standard policy that every survey project has to be metadata-driven: all statistical activities in the Agency abide by these metadata to ensure data accuracy and consistency.

The metadata has to be managed through its entire lifecycle, as in a data lifecycle, from metadata creation to metadata retirement or archiving. The metadata database has to be constantly updated so that each new survey will refer to the most current metadata.

## Key themes and long run goals

### Strategy

- NSI processes for metadata collection, migration strategy from existing fragmented metadata environments

### Integration

- international standards, local metadata needs, mapping in the NSI metadata repository

### Exchange

- International organizations → metadata obligations to be fulfilled while minimising the internal reporting burden

### Dissemination

- Publication of the metadata on the NSI website and its relation to the data dissemination

# Takeaway

## Structural

DIMENSION		DIMENSION		DIMENSION		ATTRIBUTE				
Ref. Area		Indicator		Time Period		Unit Multiplier				
Country - Pays		1-1 Total mid-year population - Population totale au milieu de l'année				Thousands - milliers				
Country - Pays		1980	1985	1990	1995	1999	2000	2001	2002	2003
Angola.....		6993	8754	9194	11072	12692	13134	13533	13942	14366
Botswana.....		906	1083	1276	1487	1529	1541	1549	1552	1565
Lesotho.....		1339	1538	1792	2050	2037	2035	2050	2065	2080
Malawi.....		6183	7340	9667	11129	11270	11308	11554	11806	12064
Mauritius - Maurice.....		966	1020	1057	1117	1151	1161	1169	1178	1187
Mozambique.....		12095	13711	14187	16004	17808	18292	18616	18946	19283
Namibia - Namibie.....		1030	1518	1349	1540	1711	1757	1787	1817	1848



### Occupancy of tourist accommodation establishments (tour\_occ)

Reference Metadata in Euro SDMX Metadata Structure (ESMS)  
 Compiling agency: Eurostat, the statistical office of the European Union

Eurostat metadata
Reference metadata
<ol style="list-style-type: none"> <li>1. Contact</li> <li>2. Metadata update</li> <li>3. Statistical presentation</li> <li>4. Unit of measure</li> <li>5. Reference Period</li> <li>6. Institutional Mandate</li> <li>7. Confidentiality</li> <li>8. Release policy</li> <li>9. Frequency of dissemination</li> <li>10. Accessibility and clarity</li> <li>11. Quality management</li> <li>12. Relevance</li> <li>13. Accuracy</li> <li>14. Timeliness and punctuality</li> <li>15. Coherence and comparability</li> <li>16. Cost and Burden</li> <li>17. Data revision</li> <li>18. Statistical processing</li> <li>19. Comment</li> </ol>
Related Metadata
Annexes (including footnotes)

The empty template is the key object

National metadata			
National reference metadata			
National metadata produced by countries and release			
<a href="#">Belgium</a>	<a href="#">Bulgaria</a>	<a href="#">Czech Republic</a>	<a href="#">De</a>
<a href="#">Germany</a>	<a href="#">Estonia</a>	<a href="#">Ireland</a>	<a href="#">§</a>
<a href="#">France</a>	<a href="#">Croatia</a>	<a href="#">Italy</a>	<a href="#">L</a>
<a href="#">Lithuania</a>	<a href="#">Luxembourg</a>	<a href="#">Hungary</a>	<a href="#">P</a>
<a href="#">Netherlands</a>	<a href="#">Austria</a>	<a href="#">Poland</a>	<a href="#">Pc</a>
<a href="#">Romania</a>	<a href="#">Slovenia</a>	<a href="#">Slovakia</a>	<a href="#">Ei</a>
<a href="#">Liechtenstein</a>	<a href="#">Switzerland</a>	<a href="#">Montenegro</a>	<a href="#">§</a>

## Reference



Andrea Bruni  
anbruni@istat.it