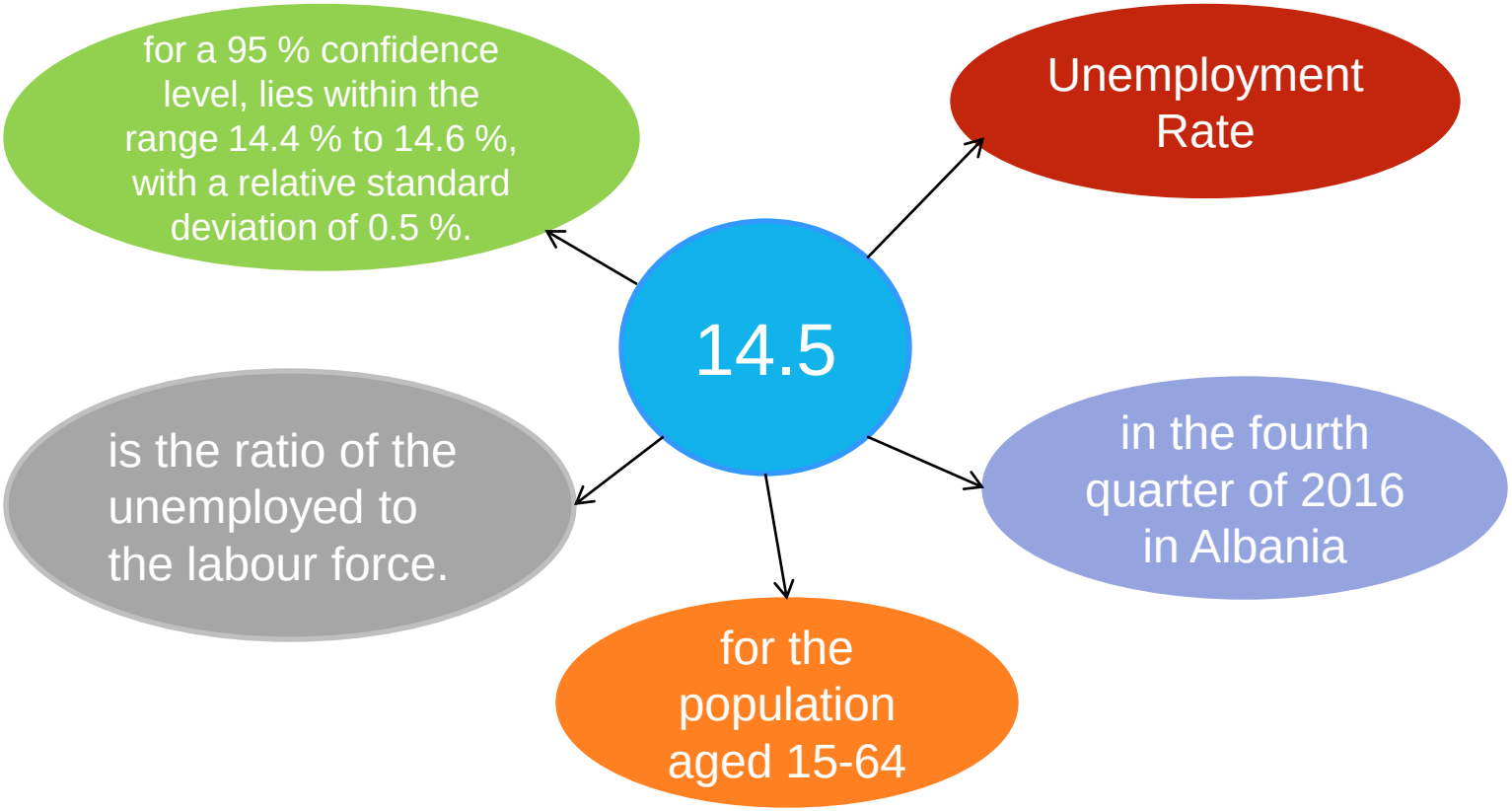
INSTAT
Republika e Shqipërisë
Instituti i Statistikave

# Metadata, the DNA of statistical data

Olta Kodra , Ertugrela Curumi, Ilda Shabani
*Institute of Statistics, INSTAT*

*International Conference on Applied Statistics and Econometrics*
*27 April 2017*

# Metadata make data significant



# Content

I. Literature review
II. Importance of Metadata in Statistics
III. Metadata and statistical data quality
    - Data Quality dimensions
    - Metadata quality dimension
IV. Managing Metadata in INSTAT
V. Finding results
VI. Conclusions

# I. Literature review

- “data about data”
- Open Forum on Metadata members concluded in: “Statistical metadata describes or documents statistical data, i.e. microdata, macrodata, or other metadata. Statistical metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data.”
- Currently EUROSTAT includes in what is called metadata:
  - Euro-SDMX Metadata Structure (ESMS)
  - International statistical classifications and nomenclatures;
  - EU legal acts and methodological manuals relating to statistics;
  - CODED (EUROSTAT's Concepts and Definitions Database)
  - Standard cross-domain code lists

# II. Metadata is important in statistics because...

*Who?*
*What?*
*How?*
*When?*
*Where?*
*Why?*

**Make information usable**

- Support consistent use of terminology
- Describe how statistical data is created, combined and shared

**Optimizes the value of existing statistical information**

- Enables broad reuse of information
- Establishes knowledge base of information based on time series metadata

**Enforces and Protects Institution reputation**

- Prevents inappropriate exposure and use of statistical data
- Provides “evidence” for governance and compliance purposes

**Minimizes costs**

- Reduces need for users to question information
- Minimizes time spent searching “inconsistent” information
- Reduces storage cost because improve reusability of information.

# III. Metadata and statistical data quality

- What defines statistics with good quality?
- ISO standard 9000:2005: "degree to which a set of inherent characteristics fulfils requirements“
- Focus user needs and how much relevant are statistical data for user needs.

# Data Quality dimensions

*Relevance* - Identify users and their expectations.

- *e.g.:* Labour Market Statistics have potential users: Ministries, National Account, Researchers. Their needs: Statistics on people in employment or not, job seekers, hours worked, income from work etc.

*Accuracy* - Degree of closeness of computations to the exact values that the statistics were intended to measure.

*Timeliness* - Length of time between data being made available and the event they describe.

- *e.g.:* Monthly External Trade data in INSTAT are published 17 days after the reference period

*Accessibility* - Conditions and modalities, by which users can obtain, use and interpret data.

- *e.g.: In INSTAT published data are available through news releases, dedicated publications and online statistical database.*

# Quality dimensions

*Clarity* - Extent in which statistical data are understandable by users

- *e.g.: INSTAT includes in the end of press releases and publications short explanations related to the definitions of the main concepts and methodological explanations*

*Coherence* - Degree to which the data and information from a single statistical programme are brought together with other data and information, and how they are logically connected and completed

- e.g.: In labour market statistics, Quarterly and Annual data are coherent and reconcilable.

*Comparability* - Extent to which statistics for a given characteristic enable reliable comparisons of values across geography and over time

- *e.g.:* CPI data in Albania are comparable between prefectures because the method used for collecting, processing and calculating is the same in all the territory of Albania.

# Metadata quality dimensions

| | |
|---|---|
| Accuracy | • Are the characteristics of the resource correctly reflected? |
| Availability | • Are metadata accessed now and over time into the future? |
| Completeness | • Are all relevant characteristics of the resource captured? |
| Credibility | • Is the metadata based on reliable sources? |
| Consistency | • Does the data not contain contradictions? |
| Process-ability | • Is the metadata properly machine-readable? |

# IV. Managing metadata in INSTAT

Users of statistical data in Albania

- Government
- Companies
- Organizations
- Researchers, journalists and students
- General public

# IV. Managing metadata in INSTAT

## INSTAT Statistical Data

### Structural Metadata

Acts as identifiers and descriptor of data

### Reference Metadata

Metadata describing the content and quality of statistical data

#### Conceptual Metadata

Describe their concept used and their practical implementation, allowing users to understand what the statistics are measuring

#### Methodological Metadata

Describes methods used for the generation of the data

#### Quality Metadata

Describe the different quality dimensions of the resulting statistics

# Structural metadata

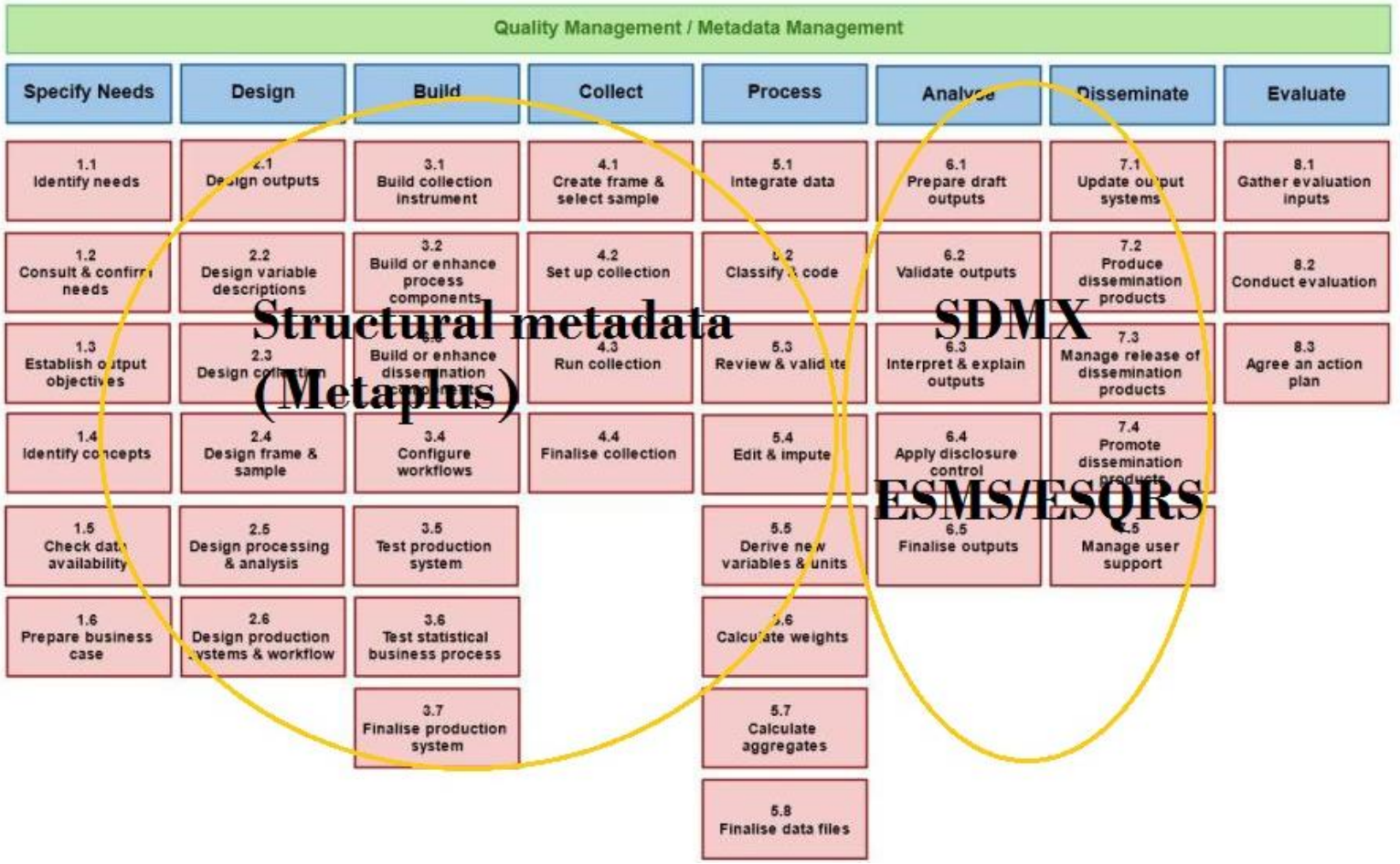
MetaPlus

- A system implemented in INSTAT in cooperation with statistics Sweden
- Documenting statistical activities and administrative datasets
- Final observation registers, describing the micro data
- Aim the unification of variables and reuse of metadata

# Reference metadata

- In 2017 INSTAT is publishing referential metadata in a EUROSTAT standard called ESMS (Euro SDMX Metadata Structure).
- Detailed standard providing considerable amount of requests.
- International standard for exchange and comparisons of statistical data in EU level.
- A desktop application is being developed in INSTAT to manage these metadata.
- The system will store, manage, publish and send metadata to EUROSTAT

# Metadata management standards and systems



# V. Results/Findings

- Invocation for metadata grows with an increasing demand on statistics and analyses based on different sets of statistics.

- Metadata has a core role in the management of data quality and it is an important component of overall management of a statistical institution.

- Data quality and metadata quality should have the same degree of importance in a statistical institution.

# VI. Conclusions

- Statistical metadata are needed to help a human user to transform statistical data into information

- Efforts on harmonisation of metadata in various processes and standards and connect these standards with each-other

- Developing and managing a statistical metadata system a concern that should involve the whole statistical institution.

# Thank You For Your Attention!