Integrated Metadata System in the Hungarian Central Statistical Office

Session number: 24
Date: 2 June 2016

Ms. Ágnes Almádi
Hungarian Central Statistical Office
agnes.almadi@ksh.hu

Madrid, May 31 - June 3

Establishing the foundation

- Research to integrate the information system with metadata
- Establishment of organisational unit → production database
- Metadata stored in MARK IV files
- IBM mainframe and batch process
- Uploading problems → restructuring
- Naming convention → ensuring connection

In the 1970s

The first metadata-driven system

- interactive accessibility was needed to aggregate on data after choosing the right metadata → SOLAR
- IBM framework on terminals → finishing the development
- Users feedback was positive → metadata-driven system was established

In the 1980s

Database management system

- Renewing the IT system → HP Unix OP system, ORACLE database management system, PC clients
- Migration of data and the programming
  - MARK IV files were migrated → new applications
  - SOLAR was not migrated → new system

In the 1990s

Metadata-driven, integrated statistical system

The metadata system is a sub-system of the statistical information system, which aims are:

- to give information on the content and quality of data or on the methods of data production to the users
- to support, document the work of persons engaged in data processing
- to support automation and integration of statistical data production and operation

Seven subsystems and their development plans [1]

1. Statistical domains

- reconsidering the List of Statistical Domains from the point of view of the whole National Statistical System
- description of the statistical domains will be upgraded according to the Single Integrated Metadata Structure (SIMS 2.0)

2. Data sources

- the description of the data sources will be updated using the widespread international standard of the Data Documentation Initiative (DDI 2.5)

3. Legal base

- in depth review and expansion will be carried out

Seven subsystems and their development plans [2]

4. Concepts

- supervision about consistency and cross-references
- will clarify key definitions and concepts for official statistics

5. Nomenclatures and classifications

- Nomenclatures:
  - setting principles for establishing new nomenclatures
  - driving factors for processing systems
- Classifications:
  - the last improvement is connected to the revision of NACE, CPA in 2008 → new database
  - changes are easy to follow up

Seven subsystems and their development plans [3]

6. Measures

- further integration → operations and connections between varieties of measures
- more measures should be visible on the website

7. Statistical registers

- the description of the ‘statistical registers’ will be also updated using the widespread international standard of the Data Documentation Initiative (DDI 2.5)
- national statistical law is expected to change → clarify the concept of ‘statistical register’ and will declare which metadata has to be published

Statistical production and metadata management

Metadata management means the description of the statistical data and other outputs and provide parameters for the control of process phases and sub-processes.
→ Overarching element of the Hungarian adaptation of the GSBPM (called ESTFM)
→ Metadata management and the IMS became strategic issue for the model

Q 2016
European Conference on Quality in Official Statistics

Quality guidelines [1]

1. Metadata must be adjusted to user needs.
2. Metadata must be timely and topical.
3. Metadata must be available for users.
4. Metadata must be comparable across the various topics and with the various time series.
5. The consistency of metadata must be ensured.
6. When metadata are generated, international standards must be followed.
7. Standard metadata must be generated and used within the HCSO.
8. Metadata must be brought in line with the various topics and time series.

Q 2016
European Conference on Quality in Official Statistics

Quality guidelines [2]

9. The completeness of metadata must be ensured.
10. Efforts must be made on ensuring integrity.
11. Metadata must be comprehensible.
12. If metadata use codes, an explanation must be provided for them.
13. Metadata must be documented.
14. Metadata must be identified and named separately.
15. Responsible persons must be allocated to metadata.
16. Metadata must be stable.
17. The necessary knowledge must be provided for the users of metadata.

Q 2016
European Conference on
Quality in Official Statistics

Thank you for your attention!