

Metadata Studies in TurkStat

TurkStat

Metadata and Standards Department

Contents

Metadata Studies in TurkStat

- ❖ Current Metadata Components
- ❖ Current Situation on Concepts Inventory
- ❖ Conclusion



Current Metadata Components

TurkStat metadata system consists of various independent components. These components are listed below:

TurkStat Web Content Management System	Process Metadata
Institutional Quality Report Form	Production Process Calendar
Classifications Server	National Data Release Calendar
Metadata Editor and Harzemli	Field Checklist for Surveys
National Data Archive (NADA)	Official Statistics Portal

Current Metadata Components-Cont'd

TurkStat Web Content Management System: Reference metadata for statistical data is disseminated in SDDS (Special Data Dissemination Standard) format in [TurkStat's web site](#).

Subject matter domains upload their data tables and metadata using the Web Content Management System.

Institutional Quality Report Form: Quality reports are available for all statistics. These reports are available on [TurkStat's web site](#).



Current Metadata Components-Cont'd

Classifications Server: National and international standard classifications and internal standard code lists are managed and accessed through the [Classification Server](#) in TurkStat's web page.

Metadata Editor: Reference and structural metadata for surveys are entered into the metadata editor in DDI 1.2 format.



Current Metadata Components-Cont'd

Harzemli: Harzemli system uses structural metadata (classifications, code lists and variables) from metadata editor to generate electronic data entry forms for surveys.

Harzemli is not only a data entry platform, it has functionalities for data editing, cleaning and processing as well.

National Data Archive (NADA): NADA is used to publish reference and structural metadata information about the surveys, which are entered using the metadata editor in DDI format.

Currently NADA is open to internal users only.



Current Metadata Components-Cont'd

Process Metadata: Business processes are modelled using ARIS. Statistical processes are modelled using GSBPM and non-statistical processes are modelled with [APQC PCF](#) (American Productivity&Quality Center Process Classification Framework) and matched with the first level of GAMS0.



Current Metadata Components-Cont'd

Production Process Calendar (PPC): PPC is used to organize and manage the statistical production process, with a focus on data collection process.

PPC allocates the workload throughout the year, specifies who (what department or unit) is responsible for what task and when the deadlines are for the tasks.

The calendar is accessible through the intranet which is open to central and regional units.



Current Metadata Components-Cont'd

National Data Release Calendar: Publication dates and times of official statistics are announced in advance by National Data Release Calendar.

This calendar helps to provide the timely release of official statistics by all institutions within the scope of the Official Statistics Programme.

[The National Data Release Calendar](#) has been prepared since 2007 by Turkish Statistical Institute in coordination with other institutions, and the calendar is updated on an annual basis.



Current Metadata Components-Cont'd

Field Checklist for Surveys: The main objectives of the Checklist are;

- to make systematic quality control of the field work,
- to identify the problems and deficiencies encountered during data collection by using built-in reports,
- to take necessary measures to overcome the problems,
- to take improvement actions for the following surveys

Official Statistics Portal: The official statistics disseminated by TurkStat as well as other government institutions are accessible through this [portal](#).



Current Situation on Metadata Exchange

TurkStat has been using Metadata Handler to send reference metadata (Euro-SDMX Metadata Structure (ESMS) and ESS Standard for Quality Reports Structure (ESQRS) files) to Eurostat for a while.

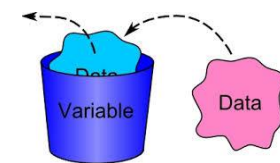
TurkStat's subject matter domains sent ESMS and/or ESQRS reports using Metadata Handler since 21/02/2012.



Ongoing Works on Concepts Inventory

A concept is a unit of knowledge created by a unique combination of characteristics.

- **A term** is a word or phrase used to designate a concept. A term is a statistical concept used in everyday statistical work. They can be found in legal acts, methodological manuals, gentlemen's agreements, etc.
- **A variable** is a characteristic of a unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned. Income, age, weight, “occupation”, “industry”, “disease”, etc.
- **Category** is a generic term for items at any level within a classification, typically tabulation categories, sections, subsections, classes and subclasses. Male and female etc.



Ongoing Works on Concepts Inventory

Stages of Concepts Inventory: Preparation

- Concepts used during the statistical production processes were put together.
- The concepts inventory includes terms, variables and categories used by Subject Matter Domains.
- The concepts inventory has created to provide content to the Concepts and Definitions Module.
- These concepts were collected by
 - Disseminated metadata by data tables
 - DDI Editor
 - Institutional databases (archives)
 - Statistical units
 - Symbols



Ongoing Works on Concepts Inventory

Draft Concepts Inventory include

	Number of Rows
Terms	84
Statistical Units	13
Symbols	24
Used Variables in Web	7913
Used Variables in DDI	553
Used Variables in Institutional Databases	1747
Indicators	149
Total	10483

Ongoing Works on Concepts Inventory

Stages of Concepts Inventory: Collection

- In this stage, we worked closely together with Subject Matter Domains (SMDs).
 - SMDs associated with concepts and statistics name by their statistics.
 - SMDs added on new and missing concepts for the concepts inventory.
 - SMDs added definitions of the new concepts.
 - SMDs offered new concepts name instead of old or unused concepts.



Ongoing Works on Concepts Inventory

Stages of Concepts Inventory: Aggregation

Different excel worksheets were combined into one.

- Files from the Subject Matter Domains were combined into a single excel file, and a combined concept inventory file was created, in which all concepts were associated with statistics.
- If same concept is used for more than one statistics, duplication studies was done for these concepts.
- Unused concepts were determined.
- Concepts inventory was created for Concepts and Definitions Module in TurkStat Metadata Management System.



Conclusion

- For the standardisation of these concepts, working group meetings should be made with related units.
- Related groups should be work together with other groups for to use same concepts for same meaning.
- The work of standardisation of the concepts will be a long-term project.



Thank you...

