

# Case study: Turkish Statistical Institute

## Case study: Turkish Statistical Institute Turkish Statistical Institute: use of GSBPM

|                 |   |
|-----------------|---|
| Contact person* | Nilgün Dorsan                             |
| Job title       | Head of Metadata and Standards Department |
| Email           | nilgun.dorsan@tuik.gov.tr                 |
| Telephone       | +903124100267                             |

### Summary\*

A brief overview of TurkStat's metadata system is provided in the case study. Usage of GSBPM for TurkStat's processes is also briefly explained.

## Metadata strategy

Turkstat developed a strategy to create a metadata system that is comparable with international standards. The aim is to integrate the metadata stored in various places into one system; so that once metadata is created and stored, it can be reused everytime it is needed. The metadata is planned to be stored in a flexible, modular system so the system can be expanded and adapted according to the changing user needs and organizational needs. The system need to have the functionality of producing metadata reports comparable with different international standard formats.

## Current situation

Currently the metadata system consists of the reference metadata published on the website and the metadata stored in production databases in DDI format. In addition to these two main systems, there are also other metadata kept in isolated systems.

Reference metadata about statistics are published in Turkstat's web page in SDDS format. Definitions of variables and concepts about statistical data are published in the web page for each statistics/product separately. The classification server is accessible through the web. It contains the national and international classifications (<http://tuikapp.tuik.gov.tr/DIESS/ChangeLocaleAction.do?dil=en>).

In 2012 TurkStat began to use a DDI tool to catalogue metadata. This new tool holds the reference metadata for surveys along with the structural metadata. The structural metadata (datasets and variable groups) is used as an input by the Harzemli system to create electronic survey forms. Therefore the data structure defined in the metadata editor is directly used in one of the components of the production system.

A comprehensive metadata management system is designed to cover the components of both ESMS and ESQRS, as well as all other metadata that is produced within the organization stored in various places and managed by using different tools. The new system is planned to be built in phases, and planned to provide a link among different sources of metadata.

## Metadata Classification

**Structural metadata:** This type of metadata is used to define the data structures. Variable names, classifications, standard code lists, variable types, data set definitions are parts of the structural metadata.

**Reference metadata:** The metadata that describes the content and quality of statistical data. The aim and scope of the study, data collection and processing methods, quality indicators are parts of the reference metadata.

## Metadata system(s)

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

**TurkStat Web Content Management System:** Reference metadata for statistical data is disseminated in SDDS format in TurkStat's web site.

**Instituttional Quality Report Module:** Quality reports are available for selected statistics. These reports are available to internal users.

**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. <http://tuikapp.tuik.gov.tr/DIESS/ChangeLocaleAction.do?dil=en>

**Metadata Editor and Harzemli:** Reference and structural metadata for surveys are entered into the metadata editor in DDI format. Harzemli system uses structural metadata 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.

**NADA:** NADA is used to publish reference and structural metadata information about the surveys, which are entered using the metadata editor in DDI format. Also there are additional methodology and other documentation made available through NADA. Currently NADA is open to internal users only.

**Field Checklist for Surveys:** The main objectives of the Checklist are; i) to make systematic quality control of the field work, ii) to identify the problems and deficiencies encountered during data collection by using built-in reports, iii) to take necessary measures to overcome the problems and to take improvement actions for the following surveys.

**Process Metadata:** Business processes are being modelled using ARIS. Statistical processes are modelled using GSBPM and non-statistical processes are going to be modelled using GAMSQ.

**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.

**Document Repository:** Documentation of surveys, statistics, projects and the other main tasks carried out within the organization.. A repository to publish these documents is a work in progress.

**Revisions:** Revisions that have been done in the past 25 years are published in TurkStat's web site for each statistics.

**Official Statistics Portal:** The official statistics disseminated by TurkStat as well as other government institutions are accessible through this portal. <http://www.officialstatistics.gov.tr/?q=en>

**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.

## Costs and Benefits

The metadata editor (DDI tool) provide the basis for electronic data entry forms for surveys. Harzemli system which is essentially data collection, data editing and cleaning, some parts of data processing are done within the new tool called "Harzemli". With the use of this tool the timeliness and effectiveness of data collection process is improved. Data dissemination and delivery times moved up.

## Implementation strategy

The metadata system is designed to include all pieces of metadata created in the organization. Building and implementation of the system is done in phases. The system is designed in a modular structure, so the modules that have priority can be built first. The parts of the system is built piece by piece and can be expanded whenever the new metadata needs arise.

## IT Architecture

## Metadata Management Tools

Metadata management tools include the tools and systems listed under Metadata Systems section of the case study. The main issue with these tools is that the metadata kept in isolated systems in the organization. This creates problems when maintaining and updating the information within these systems. Also it brings extra burden to subjects matter departments. Since there is not a comprehensive and expandible metadata system, when new needs arise the current metadata systems may not allow us to integrate these new metadata.

## Standards and formats

TurkStat currently uses SDDS structure for dissemination and DDI standards for production metadata. There is a plan to use SDMX in the future. Right now SDMX is only used to exchange metadata with Eurostat via Metadata Handler.

## Version control and revisions

Metadata in DDI format are updated on an annual basis for the surveys. Previous versions are kept in a repository.

## Outsourcing versus in-house development

Most components of the metadata systems are developed in house by TurkStat IT experts.

## Sharing software components of tools

## Overview of roles and responsibilities

The responsibility of metadata was dispersed until recently. Subject matter departments and IT department were responsible for metadata systems and metadata standards. In 2012 a new department, Metadata and Standards Department, was established. The new department is responsible for coordinating the metadata, classifications, standardisation and process modelling works in TurkStat. Subject matter departments are currently responsible for providing metadata in web site. Metadata in production systems (DDI) is created by subject matter departments and controlled by Metadata and Standards Department in terms of completeness and compliance with the standards.

Adaptation of international classifications, preparation of correspondence tables and dictionnaires for standard classifications are also done by Metadata and Standards Department. The classifications are uploaded to the Classification Server and during the built phase of the surveys, all subject matter departments are directed to use these standard code lists and international/national classification from this server. The Classification Server acts as a custodian not giving permission to use non accepted standards.

## Metadata management team

## Training and knowledge management

TurkStat's Training and Research Center arranges training programs about metadata, standards and classifications. These training are given to organizations that are in the Official Statistical Programme and TurkStat personnel in central and regional offices. There is a forum to ask questions about the classifications and coding to provide answers and to share experiences.

## **Partnerships and cooperation**

.

## **Other issues**

.

## **Lessons learned**

There have been many efforts to improve the metadata system in the past 10 years. Within the scope of these efforts, attempts have been made to develop new systems and softwares for metadata. However these efforts did not find enough support within the organization. There was not an established metadata strategy and the targets were not clearly set. Therefore there was always a risk when developing a new metadata system, because it can be disrupted as the senior management changes. New management might have a different approach towards metadata, objectives and priorities may change and the previous works could come to a stop. Therefore it is important to have a clear strategy and targets from the beginning, and organizations need to consistently work towards reaching these targets. Senior management's support as well as organizational support is extremely important to reach these targets.

Metadata works are usually not supported by the departments who enter and update the metadata information in the systems as well because of the burden of entering and updating the metadata. Organization wide awareness about the need for an established metadata system did not exist until recently. But this point of view is changing. Transition period might be burdensome, but a good system is useful for everyone and once it is established there will be less work. The subject matter departments are starting to support the change after seeing the benefits of the new systems and tools.

|               |
|---------------|
| <b>Links:</b> |
|               |

## **Attachments**