Regional Training on "Metadata Preparation Management" in cooperation with Turkish Statistical Institute - SESRIC

9-11 September 2013
Abu Dhabi, United Arab Emirates

Nalan EVİREN
TURKSTAT Expert

What is Process ?

A process is made up of a set of activities that transform inputs into outputs in such a way that added value is generated. For a process to be considered as such, it needs to have resources, which are the elements that allow the process to be performed.

A process “interacts” with various interested parties (customers of the process, people in the organization and society), by influencing the quality of the product, the health and safety of people and environmental aspects.

What is Process ?

A process has some sub-processes. And each of these sub-processes may have some sub-sub-processes.
According to process modelling theory, each sub-process should have a number of clearly identified attributes, including:

· Input(s);
· Output(s);
· Purpose (value added);
· Owner;
· Guides (for example manuals and documentation);
· Enablers (people and systems);
· Feedback loops or mechanisms.

What is Process Management?

Administrative activities aimed at
(1) defining a process, (2)
(2) establishing responsibilities,
(3) evaluating process performance,
(4) and (4) identifying opportunities for improvement.

Process-based management emphasizes monitoring and measurement to assure that the results expected in these areas are obtained.

What is Generic Statistical Business Process Model (GSBPM)?

The GSBPM is a flexible and reference tool to describe and define the set of business processes needed to produce official statistics.

What is Generic Statistical Business Process Model (GSBPM)?

The original intention is for the GSBPM to provide a basis for statistical organizations ***to agree on a standard terminology*** to aid their discussions on developing **statistical metadata systems and processes**.

A generic model was decided for statistical business processes which was named Generic Statistical Business Process Model (GSBPM), at the end of the work sessions of the Joint UNECE / Eurostat / OECD on Statistical Metadata in 2009,

Why GSBPM is needed?

- To define and describe statistical processes in a coherent way
- To standardize process terminology
- To compare and benchmark processes within and between organisations
- To identify synergies between processes
- To inform decisions on systems architectures and organisation of resources

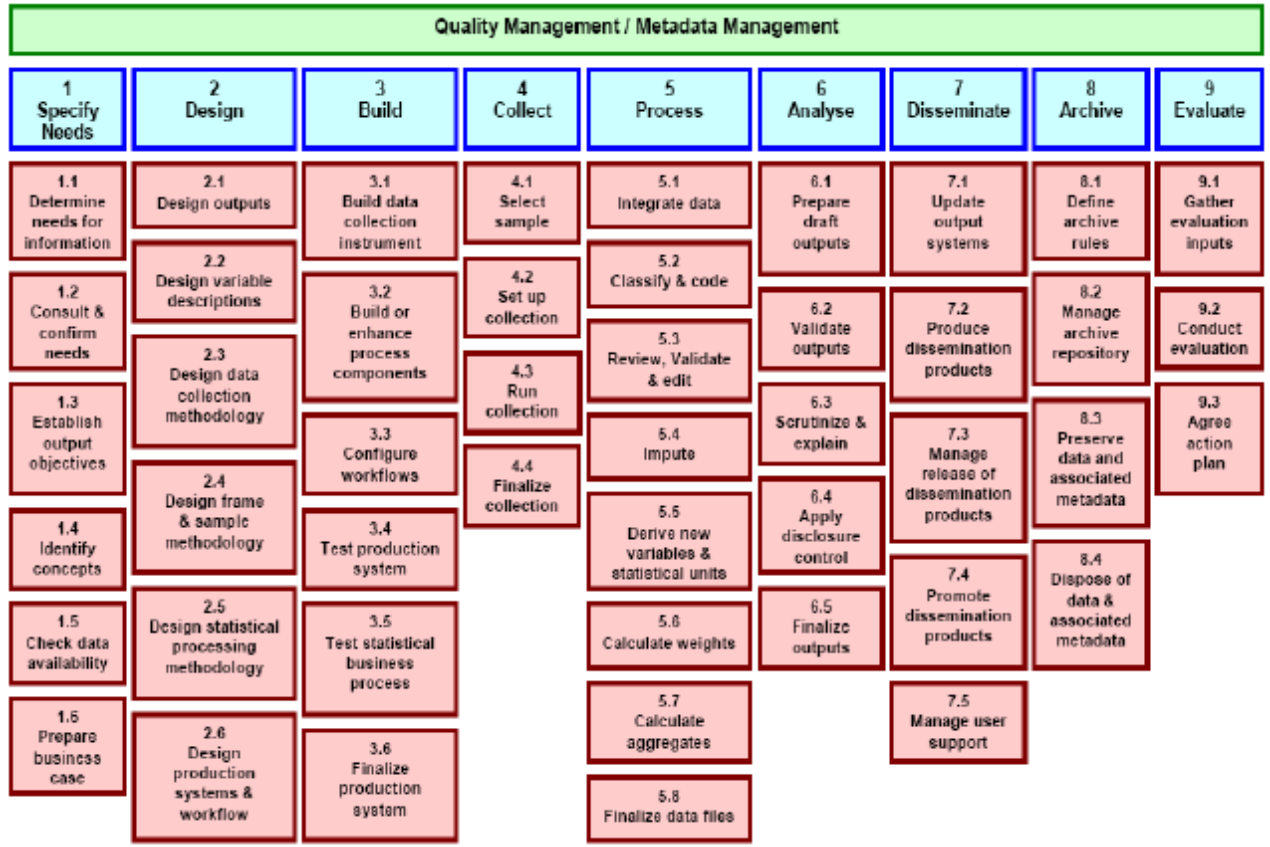
Structure of GSBPM

Process

Phases

Sub-processes

(Descriptions)



Structure of GSBPM

The GSBPM is composed of 9 phases (Each phase have some sub-processes) :

1. Specify Needs (with 6 sub-processes) :

In this phase the organization:

- determines the need for the statistics;
- confirms, in more detail, the statistical needs of the stakeholders;
- establishes the high level objectives of the statistical outputs;
- identifies the relevant concepts and variables for which data are required;
- checks if current collections and / or methodologies can meet these needs;
- prepares the business case to get approval to produce the statistics.

Structure of GSBPM

2. Design (with 6 sub-processes) :

This phase describes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes.

Structure of GSBPM

3. Build (with 6 sub-processes) :

This phase builds and tests the production systems to the point where they are ready for use in the “live” environment.

Collect (with 4 sub-processes) :

This phase collects all necessary data, using different collection modes (including
extractions from administrative and statistical registers and databases), and loads them into the appropriate data environment.

Structure of GSBPM

5. Process (with 8 sub-processes) :

This phase describes the cleaning of data records and their preparation for analysis.
It is made up of sub-processes that check, clean, and transform the collected data, and may be repeated several times. For statistical outputs produced regularly, this phase occurs in each iteration. The sub-processes in this phase can apply to data from both statistical and non-statistical sources.

Structure of GSBPM

6. Analyse (with 5 sub-processes) :

In this phase, statistics are produced, examined in detail and made ready for dissemination. This phase includes the sub-processes and activities that enable statistical analysts to understand the statistics produced. For statistical outputs produced regularly, this phase occurs in every iteration. The Analyse phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced.

Structure of GSBPM

7. Disseminate (with 5 sub-processes) :

This phase manages the release of the statistical products to customers.

8. Archive (with 4 sub-processes) :

This phase manages the archiving and disposal of statistical data and metadata.

9. Evaluate (with 3 sub-processes) :

This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general over-arching process of statistical quality management.

GSBPM and the metadata management

Metadata management is essential for the efficient operation of statistical business processes.

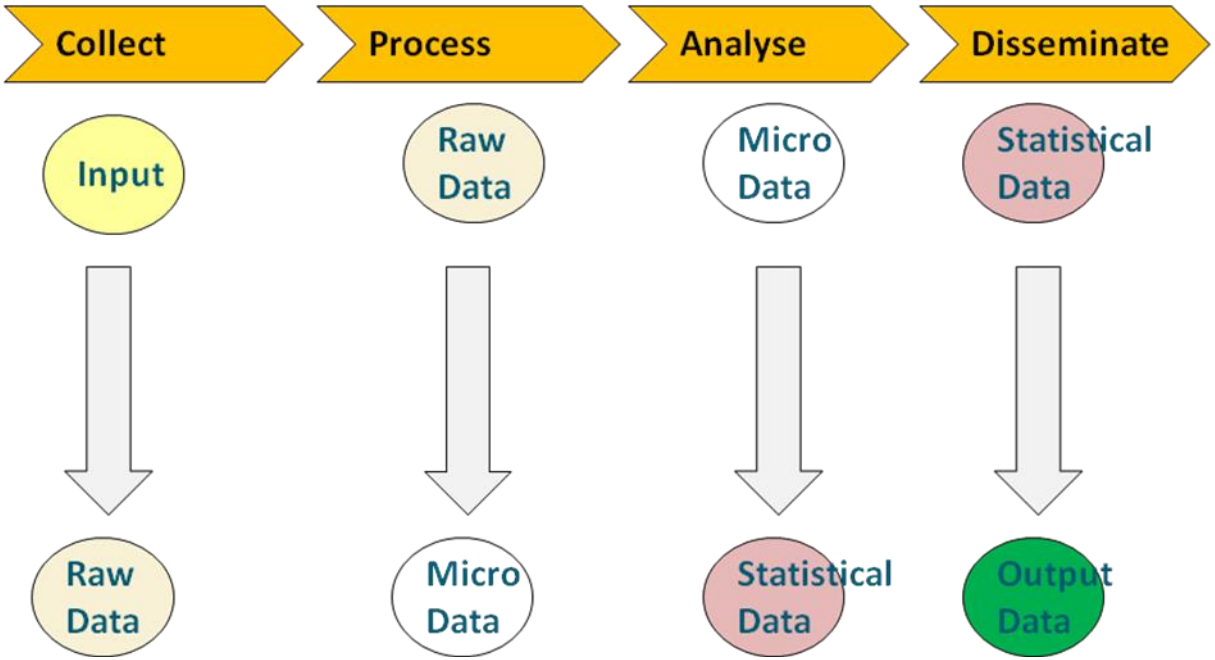
Metadata are present in every phase, either created or carried forward from a previous phase. In the context of the model, the emphasis of the over-arching process of metadata management is on the creation and use of statistical metadata, though metadata on the different sub-processes themselves are also of interest, including as an input for quality management.

GSBPM and the metadata management

The key challenge is to ensure that these metadata are captured as early as possible, and stored and transferred from phase to phase alongside the data they refer to. Metadata management strategy and systems are therefore vital to the operation of this model.

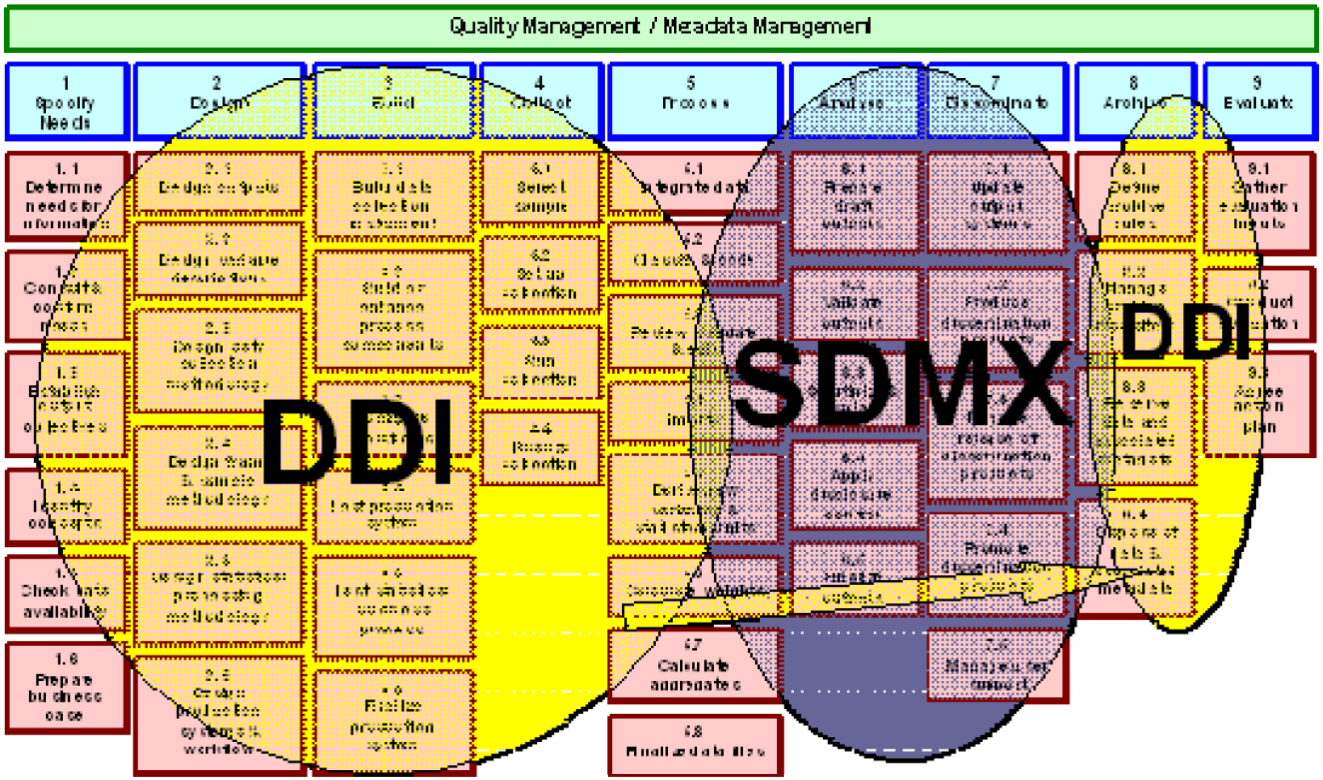
GSBPM and the metadata management

Metadata are present in every phase, either created or carried forward from a previous phase.



GSBPM and the metadata management

The relationship between DDI- SDMX and GSBPM



GSBPM and the National Statistic Offices

- Statistical agencies that have adopted GSBPM as is,
- Statistical agencies that adopted a version of GSBPM that fits with their agency,
- Statistical agencies that have an existing process model that can be mapped to GSBPM, and
- Statistical agencies that do not have an existing process model.

TurkStat Business Process Model

- Statistical production process as a whole is currently being discussed in many platforms.

- The main aim is the same for all statistical offices: the necessities of being in the information age, the need to produce consistent and relevant statistics, establishment of strong metadata system within statistical business processes.

- A metadata infrastructure which is comparable with the international standards and the reusabilty of these standards were important starting points for TurkStat.

TurkStat Business Process Model

An action plan was prepared for the standardization of data and the information by making an analysis of the needs of TurkStat. The aims of the action plan are;

- To supply the documentation in every processes
- To prepare the infrastructure for the development of metadata system
- To standardize code lists used the production process
- To make the management and follow up easy by centralizing the metadata system,
- To strengthen the institutional memory,
- To prevent reiterations in business processes and increasing the reusability of the standardized processes

TurkStat Business Process Model

GSBPM has been adopted as a reference model to bring together the metadata that get created in all phases of production.

A project started in 2010 in TurkStat to define the processes for all statistical products produced in TurkStat.

In order to specify the processes in TurkStat, four main operational processes were analysed first: collect, process, analyse, disseminate.

TurkStat Business Process Model

Five products were selected to be analysed for the pilot study of the GSBPM model.

These products were;

- short term industrial statistics,
- household budget statistics,
- crop production statistics,
- producer prices,
- consumer prices.

TurkStat Business Process Studies

Process maps and work flows for these products were drawn.

It was seen from the pilot study that the other 3 processes (specify needs, design, build) can be clearly identified too.

Based on the process flows of the pilot study a national draft model of GSBPM was developed.

This draft model was used as a basis for process modelling and standardization of the remaining products. Approximately 250 statistical products were analysed.

TurkStat Business Process Studies

Within the scope of the process modelling and standardization project,

Meetings were held with all units producing statistics to obtain information about the activities done to produce these statistics.

Procedures/methods that are used, step by step work flows, inputs and outputs of sub processes, owners of the processes, the number of people required to do these tasks, software tools and other relevant metadata were collected from the subject matter departments.
Process maps and work flows for these products were drawn.

TurkStat Business Process Studies

All the process information collected from the departments is currently being analysed, grouped and standardised.

The information objects are going to be identified.

Different procedures and algorithms used in different products are going to be listed.

The draft model will be finalized after the process analysis was completed for all products and processes.

TurkStat Business Process Studies

Draft Statistical Business Process Model

1.Specify Needs	2.Design	3.Build	4.Collect	5.Process	6.Analyse	7.Disseminate
1.1.Determine need for information	2.1.Design statistical products and outputs	3.1.Build and enhance production system components	4.1.Establish frame and registers, select sample	5.1.Clasify and code	6.1.Evaluate the information for its effect	7.1.Update dissemination systems
1.2.Consult and confirm need	2.2.Design frame, register and sample methodology	3.2.Integrate production system with other systems	4.2.Set up collection	5.2.Micro-edit	6.2.Produce statistics	7.2.Produce dissemination product
1.3.Establish output objectives	2.3.Design data collection methodology	3.3.Test production system	4.3.Run collection	5.3.Macro-control	6.3.Quality assure statistics	7.3.Manage publishing for dissemination product
1.4.Check data availabilty	2.4.Design process and analysis metodology	3.4. Finalise production system	4.4.Finalise collection	5.4.Imputation	6.4.Examine and evaluate statistics	7.4.Manage user demends
1.5.Determine business plan	2.5.Design production system and workflows			5.5.Calculate weights and derive variables	6.5.Prepare statistics for dissemination	
					6.6.Finalise content	

Draft model used in Turkstat

TurkStat Business Process Studies

The model comprises of three levels :

Level 1: seven phases of the statistical business processes as follows:

1.Specify needs,
2.Design,
3.Build,
4.Collect,
5.Process,
6.Analyse,
7.Disseminate.

TurkStat Business Process Studies

Level 2, the sub processes within each phase.

Level 3, the sub sub processes within each sub process.

Metadata are generated and processed within each phase of the model, there is, therefore, a strong requirement for a metadata management system to ensure that the appropriate metadata retain their links with data throughout the model.

TurkStat Business Process Studies

Metadata are generated and processed within each phase of the model, there is, therefore, a strong requirement for a metadata management system to ensure that the appropriate metadata retain their links with data throughout the model.

What is Metadata ?

Meta data describes an information resource. The term ***meta*** *derives from the Greek.* It means "denoting a nature of a higher order or more fundamental kind," such as *metalanguage or metatheory.*

Metadata, then, is ***data about data.***

Meta data is the meaning of the data, the quality of the source, format, and it identifies the variables. Meta data is something that converts data into information.

What is Metadata ?

A statistical data is represented as follows:

- 74.724.269
- You will have no idea of what it actually represents .
- A number of questions come to mind ???

What is Metadata ?

- What is the subject of the measurement?
- Which is measured in the unit?
- Which country or geographic region?
- When was the measurement made?
- … ?

If I tell you the answers to these questions, the data will begin to make sense

What is Metadata ?

The answers to these questions:

- Subject is “Total population.”
- Measurement unit is “Person”
- Country is “Turkey”
- Time is “31 December 2011”

Metadata

74.724.269 → Total Population of Turkey by 31 December 2011

74.724.269 →

Total population of Turkey by 31 December 2011.

Subject "Total population."
Measurement unit "person"
Country "Turkey"
Time "31 December 2011"

What is Statistical Metadata ?

Statistical metadata contains key informations about statistical information & data in the manner of meaning , quality, compliance information as well. Statistical metadata is defined in detail as a structured and systematic information used in the processes of statistical production, distrubution.

The most important point in managing all kinds of statistical processes is the efficient use of metadata management.

What is Metadata Management ?

Meta data management ensure that information needed to support management and business intelligence decision making is accurate, accessible, complete, consistent, timely, valid and relevant and assists in maintaining high integrity over production data.

Metadata management works on building productivity, enhancing data quality, cost saving on business activities and reduce the redundancies.

Metadata Management

Purpose:

Creating data structures for all statistical production processes according to international standards, and establishment of a metadata system integrated to statistical business processes.

The primary step in creating an effective system of statistical production is to design data and metadata structures, And to put into effect these structures

Metadata Management

Central meta-data structure should be defined in detail level and designed as to serve all the structures of statistical production.

In addition, this meta data structure must also ensure the connection of micro and macro data.

In summary, throughout the life cycle of all statistical production, meta-data should be identified at the produced time –begining - and used in the process if necessary.

Metadata Repository

Metadata Repository is one of the outcomes of metadata management process. Metadata management is an end-to-end process for creating, enhancing and maintaining meta-data repository and associated processes.

Metadata repository contains all details on an organization management environment. These details are placed in a central repository or in well-connected synchronized repositories. An ideal situation when there are multiple repositories, one does run into a challenge of making sure that they are well-synchronized and integrated.

Key Rules of Metadata Management

✓ The whole Statistical Business Process has to be driven by metadata.

✓ Metadata are systematically collected at their source and are subsequently reused and augmented repeatedly throughout the statistical production process.

✓ The metadata system has to include all stages of the statistical production line.

✓ Facilitate communication and **exchange of information with both national and international** bodies and end users.

Why Metadata?

Standardization

- Common Processes
- Common Concepts
- Common Variables
- Common Code List …

Institutional Memory

- Documentation
- Reuse

Why metadata ?

Data Integration and Interoperability

- Promotes interoperability and standardization across organizations
- Avoid dublication

Accessibility

- Makes data accessible to the anyone who wants to understand and use

Quality

Where Metadata?

- In which processes we need metadata?
  - Just in **Production? (Structural Metadata)**
    - Definition of Variables
    - Creation of Survey Forms
  - Just in **Dissemination? (Reference Metadata)**
    - Scope of the study
    - Concepts of study
    - Methodology used in study
    - Quality issues
  - Actually No.
    - We need metadata in every where data exists.
    - We need metadata in **all statistical processes.**

Fundamental Metadata Concepts

- Reference metadata
- Structural metadata
- Process metadata
- Quality metadata

Fundamental Metadata Concepts

✓For the description of a statistical operation reference metadata are used.

✓ For the description and handling of the databases the structural metadata are used.

✓ For the description of the process and its subprocesses the GSBPM and process metadata are used.

✓For the description of a statistical timeliness, accuracy, completeness quality metadata are used.

Fundamental Metadata Concepts

Structural Metadata

Structural meta data is metadata that is used to define the structure of the data. Variable names and potential value range, classifications, standard code lists,unit of measurement, variable types and definitions are covered by the structural metadata.

Fundamental Metadata Concepts

STANDARD CODE LIST (PRODUCTION DATABASES)

Code : EDUCATION	Description
1	Primary school
2	High school
3	Vocational or technical high school
4	Open education (primary, secondary School)
51	2 or 3-year college
52	4-year college or faculty
53	5 or 6-year faculty
61	Master
62	Doctorate

Fundamental Metadata Concepts

STANDARD CODE LIST (PRODUCTION DATABASES)

VARIABLE NAME: EDUCATION

LABEL : Completed school

DATA FORMAT: Numeric

MIN : 1

MAX: 98

Fundamental Metadata Concepts:

Reference metadata

Reference metadata that defines the content and quality of statistical data.

The purpose of the study, the scope of the data collection and processing methods, quality and delivery indicators etc. are covered by the reference metadata.

Fundamental Metadata Concepts

Reference Metadata

	Description
Title	Household Labour Force Survey,2011
ID	TR-TUIK-NDID-LFS-2011
Abstract	Household Labour Force Survey which have regularly been applied since 1988 is the main data source on the labour market situation of country from the supply side and gives information on economic activity, occupation, status in employment and hours worked for employed persons; and information on the duration of unemployment and occupation sought etc. by the unemployed.
Keywords	Labour force, Employment, Employment Rate
Data sources	Data were collected from the households which were selected by defined sampling method. Statistical unit used is “household” in labour force surveys. Demographic information (age, sex, educational status, relationship to household head) is asked to all members of the household. But, questions on labour force status are asked for persons 15 years old and over.

Fundamental Metadata Concepts:

Quality Metadata

Defined as the meta-data in which quality indicators defined for the generated statistics, like timeliness, accuracy, completeness etc.

Quality Meta-data is a kind of referenced metadata and focused on reporting of data quality.

According to the development in metadata application, over time, not only the quality of data, the quality of metadata started to gain importance.

Metadata Components

Concepts and definitions

One of the important steps to produce comparable statistics is to manage terms and concepts of variables defining a central structure according to specific standards.

Establishing concepts and definitions of the variables and entegrate to associated study/ work is important. Thus, the external users can access of to variable definitions and concepts in datasets and related metadata files. Also, inner users use standarts in production of statistics. As a result, a convenient, and sustainable system can be built up.

Metadata Components

Variable Bank and Statistical Units

One of the important components of Metadata system is variables. The first steps to ensure the statistics to be comparable is to the standardize the variables used in the production of statistics.
if variable names and variable values for categorical variables,- coding lists – can be collected in a pool of standard, tools of production can easily access to the variables

Likewise, statistical units should be aggregated in a standardized central structure.

Metadata Components

Classifications

One of the most important sub-components of Metadata management systems are statistical classifications. Statistical classification is an essential component of statistical production systems.

Classifications are used actively in all stages of the statistical production from the field works to the distribution process.Statistical classifications can be used in more than one statistical area like classifications of activities and geographical classifications. But also, it can be used specific to a single area.

Metadata Components

Classifications and Standard Code Lists

Classifications and standard code lists have a primary importance for the produced statistics to be comparable. The effective management of Classifications is depends on the two key points:

- can be easily accessed by anyone.
- published on a server open to anyone.

During the design of production and data validation tools, classification servers will be directly linked to them and the classifications used in the production server databases should be designed to use the most current form of classification.

Metadata Components

Classifications

Especially in our country, not only the internal users, as well as public institutions and organizations are using some of the classifications via their own systems be able to query the classification database server.

Classification Server:

Links for the Classification Server

TurkStat official website

http://www.tuik.gov.tr/

Classification Server website

http://tuikapp.tuik.gov.tr/DIESS/

Classification Server

Classification server serves as a information system for

- Institutional data collection and data preparation operations
- External users (public and other institutions) requirements
- Includes translations and adaptations of international classifications and also national classifications,
- Correspondence tables
- Dictionaries
- Forum

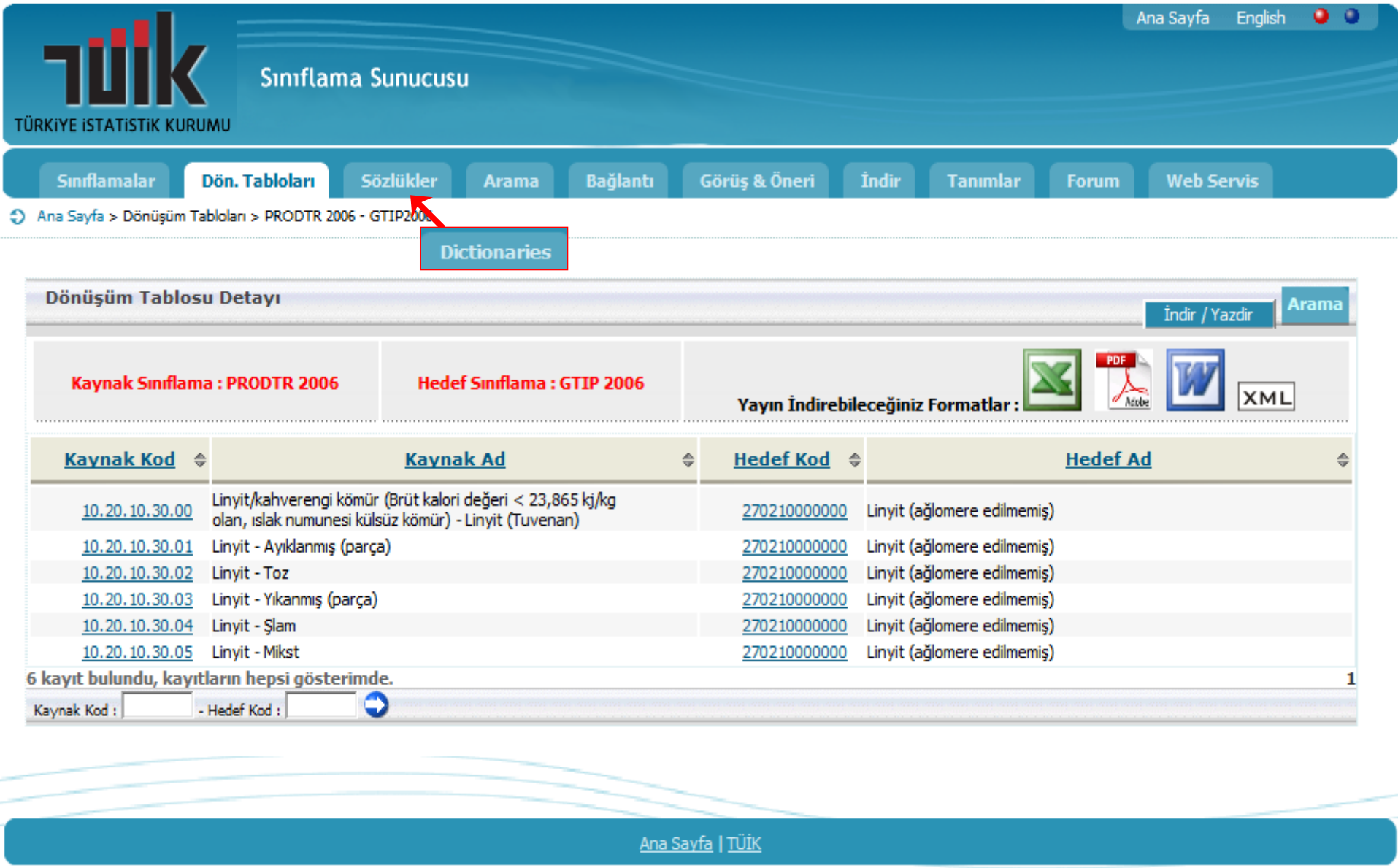
Classification Server

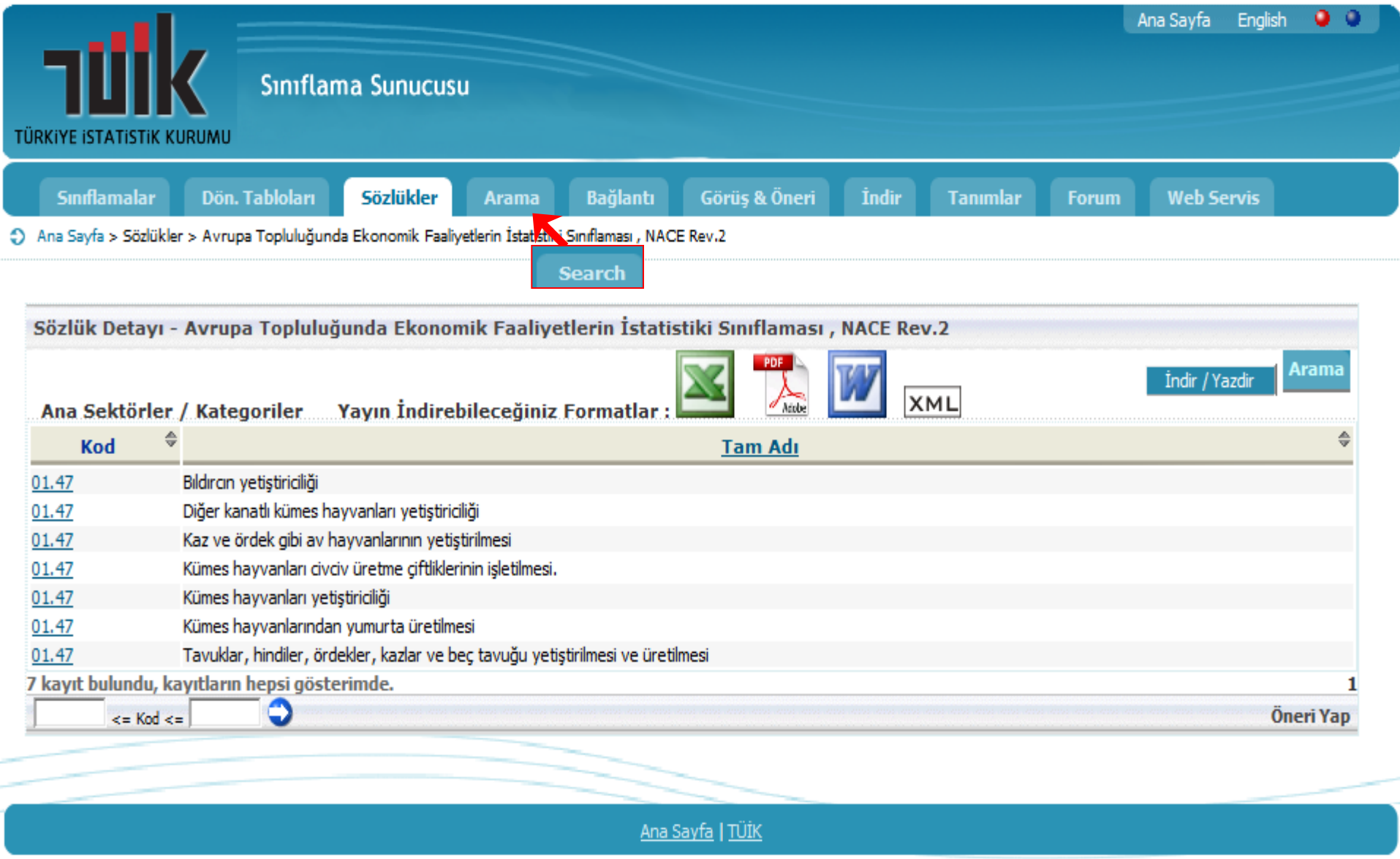
Forum

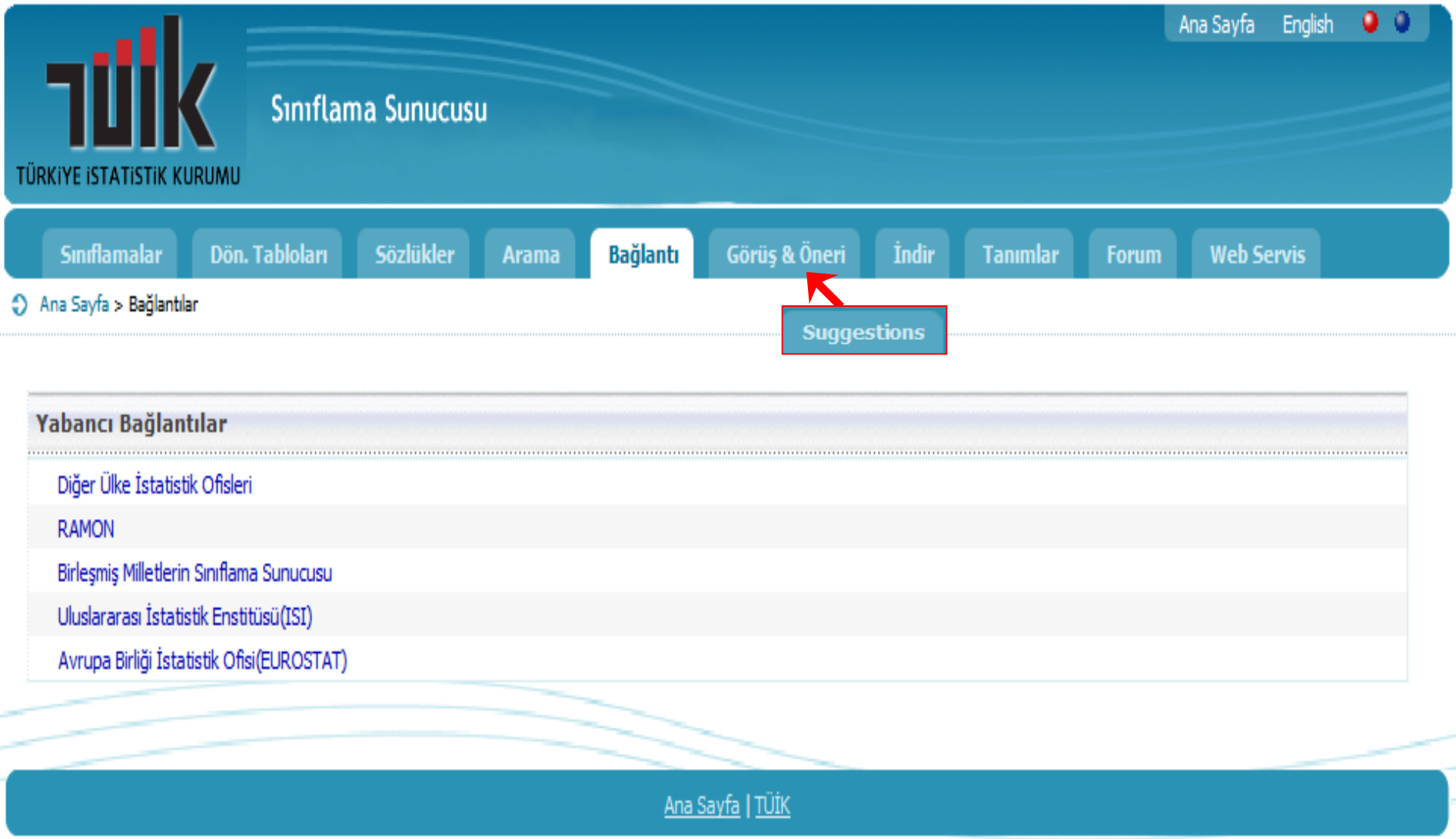
- 351 users

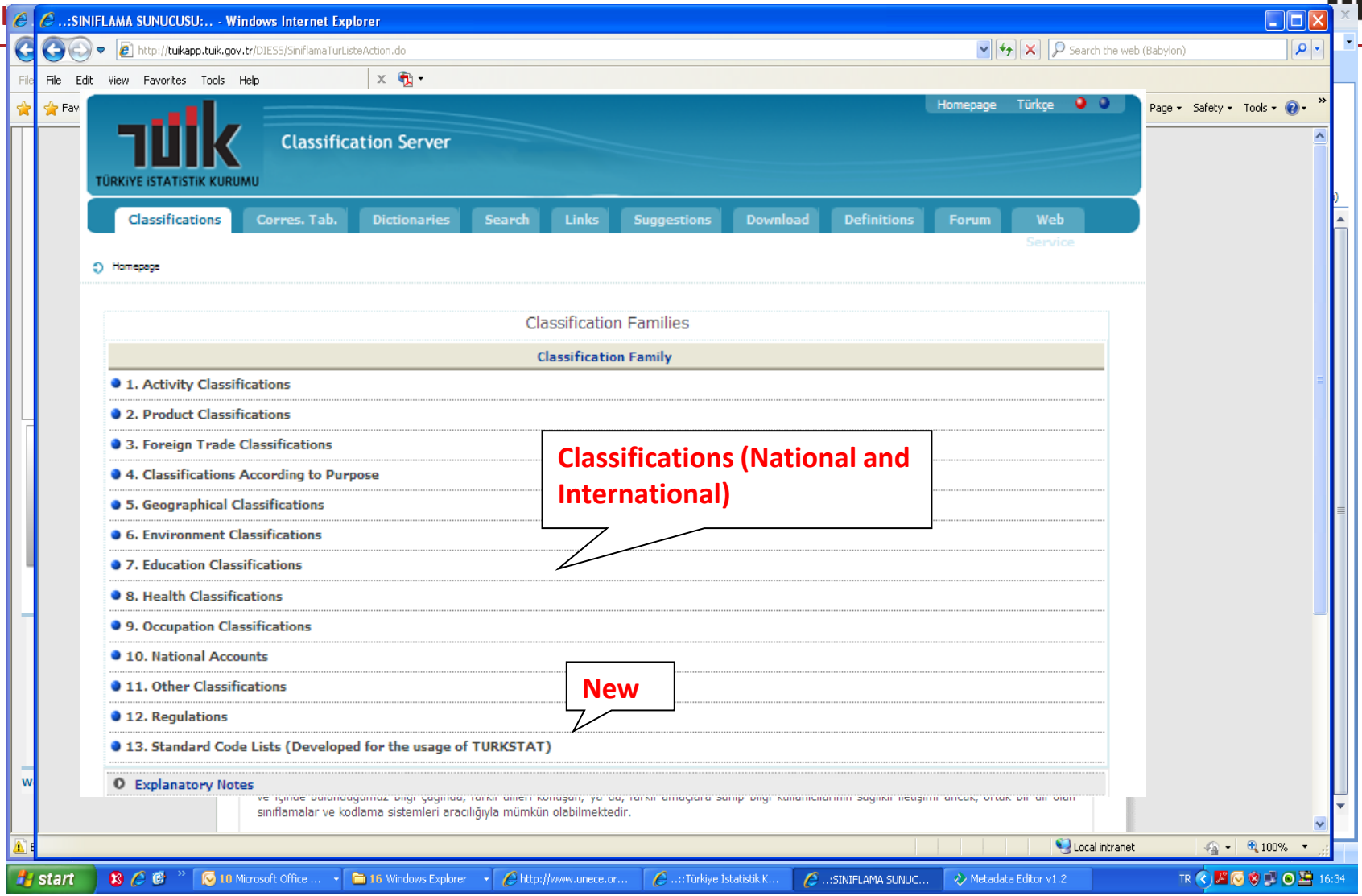
- Ministries, institutions and organizations and chambers of commerce (Ministry of Health, Ministry of National Education, some undersecretariats, Energy Market Regulatory Authority, Ankara Chamber of Commerce etc.)

- Statistical offices (Moldovia, Kazakhstan, Turkmenistan, Germany, Eurostat, Palestine etc.)


Ana Sayfa English
TÜRKİYE İSTATİSTİK KURUMU
Sınıflama Sunucusu
Sınıflamalar
Dön. Tabloları
Sözlükler
Arama
Bağlantı
Görüş & Öneri
İndir
Tanımlar
Forum
Web Servis
Ana Sayfa > Dönüşüm Tabloları > PRODTR 2006 - GTIP2006
Dictionaries
Dönüşüm Tablosu Detayı
İndir / Yazdır
Arama
Kaynak Sınıflama : PRODTR 2006
Hedef Sınıflama : GTIP 2006
Yayın İndirebileceğiniz Formatlar :
XML
Kaynak Kod
Kaynak Ad
Hedef Kod
Hedef Ad
10.20.10.30.00 Linyit/kahverengi kömür (Brüt kalori değeri < 23,865 kj/kg olan, ıslak numunesi külsüz kömür) - Linyit (Tuvenan) 270210000000 Linyit (ağlomere edilmemiş)
10.20.10.30.01 Linyit - Ayıklanmış (parça) 270210000000 Linyit (ağlomere edilmemiş)
10.20.10.30.02 Linyit - Toz 270210000000 Linyit (ağlomere edilmemiş)
10.20.10.30.03 Linyit - Yıkanmış (parça) 270210000000 Linyit (ağlomere edilmemiş)
10.20.10.30.04 Linyit - Şlam 270210000000 Linyit (ağlomere edilmemiş)
10.20.10.30.05 Linyit - Mikst 270210000000 Linyit (ağlomere edilmemiş)
6 kayıt bulundu, kayıtların hepsi gösterimde.
1
Kaynak Kod :
- Hedef Kod :
Ana Sayfa | TÜİK
Necatibey Cad. No:114 Bakanlıklar/ANKARA Tel: 0. 312. 410 0 410
(c) 2006 TÜİK

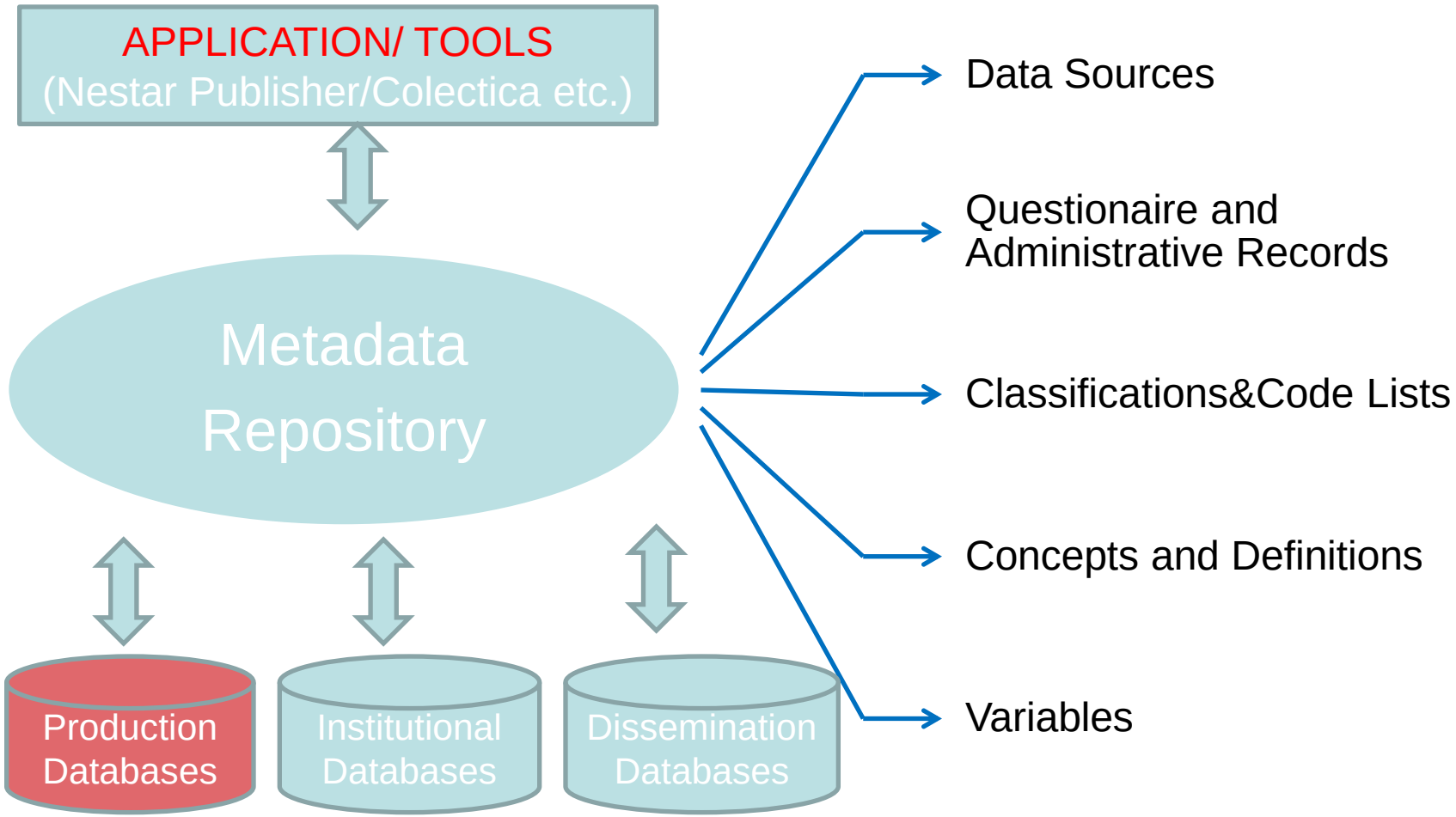

Ana Sayfa English
TÜRKİYE İSTATİSTİK KURUMU
Sınıflama Sunucusu
Sınıflamalar
Dön. Tabloları
Sözlükler
Arama
Bağlantı
Görüş & Öneri
İndir
Tanımlar
Forum
Web Servis
Ana Sayfa > Sözlükler > Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistiki Sınıflaması , NACE Rev.2
Search
Sözlük Detayı - Avrupa Topluluğunda Ekonomik Faaliyetlerin İstatistiki Sınıflaması , NACE Rev.2
Ana Sektörler / Kategoriler
Yayın İndirebileceğiniz Formatlar :
PDF
XML
İndir / Yazdir
Arama
Kod
Tam Adı
01.47 Bıldırcın yetiştiriciliği
01.47 Diğer kanatlı kümes hayvanları yetiştiriciliği
01.47 Kaz ve ördek gibi av hayvanlarının yetiştirilmesi
01.47 Kümes hayvanları civciv üretme çiftliklerinin işletilmesi.
01.47 Kümes hayvanları yetiştiriciliği
01.47 Kümes hayvanlarından yumurta üretilmesi
01.47 Tavuklar, hindiler, ördekler, kazlar ve beç tavuğu yetiştirilmesi ve üretilmesi
7 kayıt bulundu, kayıtların hepsi gösterimde.
1
<= Kod <=
Öneri Yap
Ana Sayfa | TÜİK
Necatibey Cad. No:114 Bakanlıklar/ANKARA Tel: 0. 312. 410 0 410
(c) 2006 TÜİK


Ana Sayfa
English
TÜRKİYE İSTATİSTİK KURUMU
Sınıflama Sunucusu
Sınıflamalar
Dön. Tabloları
Sözlükler
Arama
Bağlantı
Görüş & Öneri
İndir
Tanımlar
Forum
Web Servis
Ana Sayfa > Bağlantılar
Suggestions
Yabancı Bağlantılar
Diğer Ülke İstatistik Ofisleri
RAMON
Birleşmiş Milletlerin Sınıflama Sunucusu
Uluslararası İstatistik Enstitüsü(ISI)
Avrupa Birliği İstatistik Ofisi(EUROSTAT)
Ana Sayfa | TÜİK
Necatibey Cad. No:114 Bakanlıklar/ANKARA Tel: 0. 312. 410 0 410
(c) 2006 TÜİK



Classification Server, including a new section for the standard code lists

Integrated Metadata System Component



Integrated Metadata System

Key Point : Integration

Even though all sub-components that make up metadata with their own specialized functions, when designed in a way that they integrate to each other, it will ensure an understandable, standard and comparable statistical production.

Standards

When Metadata & Standard comes to mind

1. DDI (Data Documentation Initiative)
   - Main Focus: Capture metadata about micro data
2. SDMX (Statistical Data and Metadata Exchange)
   - Main Focus: Exchange of Data and Metadata between organizations
3. GSBPM (Generic Statistical Business Process Model)
   - Main Focus: Capture metadata about micro data

Metadata Standards

Metadata Standards are the standards that ensure the users as well as the owners of the data to understand the data and use the data correctly by creating a sense of common standards for data.

DDI [Data Documentation Initiative]

DDI is an initiative to create an international standard expressed in XML for describing structured micro data.

Metadata Standards

1. Version1-2.5) DDI-Codebook
2. Version 3.0- 3.1 DDI-Life Cycle

These two lines have different characteristics, but both allows to identify micro data sets, variables, code lists and questions.

We use DDI-Codebook in the context of Turkstat metadata applications.

Metadata Standards

Evolution of the versions of DDI

DDI 1: microdata surveys
DDI 2: added aggregate tabular data
DDI 3: modular, life cycle model, complex data files, comparative data files

Metadata Standards

Structure of DDI 1.0/2.0

Document description: information about the DDI document and how it was created with bibliographic citation

Study description: information about the context of the data production and distribution (creators, methodology, abstract, keywords, etc.)

Data files description: information about the data file or files (format, size, number of cases, etc.)

Variable description: information about the data items or rows and columns in a tabular data file/s

Other study materials: inline reference materials or references to external reference materials (coding schemes, thesauri, citations to publications, etc.)

DDI 1 / 2

Document Description
- Citation of the codebook document
- Guide to the codebook
- Document status
- Source for the document

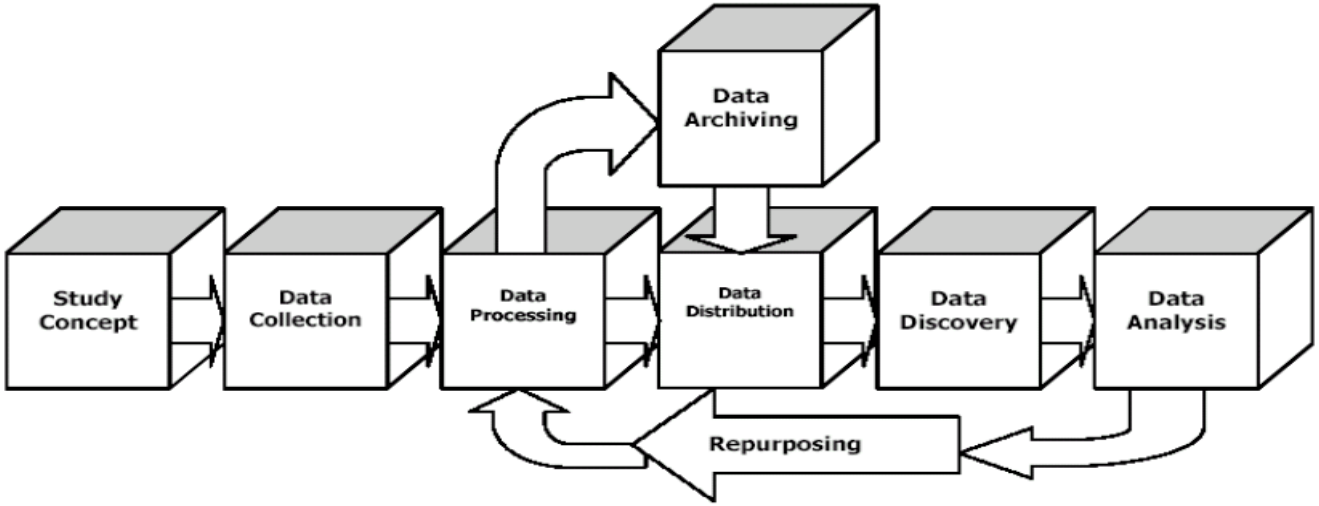
Study Description
- Citation for the study
- Study Information
- Methodology
- Data Accessibility
- Other Study Material

File Description
- File Text (record and relationship information)
- Location Map (required for nCubes optional for microdata)

Data Description
- Variable Group and nCube Group
- Variable (variable specification, physical location, question, & statistics)
- nCube

Other Material

DDI 3.0 and the Survey Life Cycle



- A survey is not a static process: It dynamically evolved across time and involves many agencies/individuals
- DDI 2.x is about archiving, DDI 3.0 across the entire “life cycle”
- 3.0 focus on metadata reuse (minimizes redundancies/discrepancies, support comparison)

XML (Extansible Markup Language)

XML-based technologies are used in today's e-government applications.

XML has become standard in order to ensure the most important characteristics such that, flexibility, extensibility, interoperability.

SDMX (Statistical Data and Metadata Exchange)

- **SDMX (Statistical Data and Metadata Exchange) is the international standards** for the electronic exchange of statistical information. It is an initiative that aims to foster common standards and guidelines for the exchange and sharing of statistical data and metadata.

- Developed for statistical tables
- Supports well structured, well defined data, particularly time-series data
- Contains both metadata and data
- Supports transfer of data between systems

SDMX (Statistical Data and Metadata eXchange)

SDMX

ISO TS 17369

World Bank

Lack of standardization of data exchange between institutions

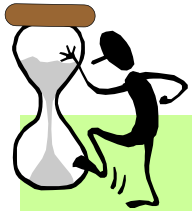
The data and the metadata in different formats

Various fields to store data and metadata

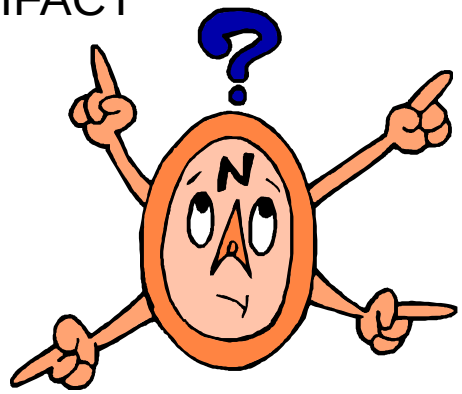


EDIFACT

XML



different tools



paper form

Email

Web-form

dial-up

Structured files

Tools(moveable)

file upload

paper

The Euro-SDMX Metadata Structure (ESMS)

The ESMS is the unique structure in use in Eurostat for the dissemination of **reference metadata** at European level. It is also the structure used for the collection of national reference metadata files from National Statistical Institutes.

SDMX Reference Meta data

Annual national accounts

Reference Metadata in Euro SDMX Metadata Structure (ESMS)

Compiling agency: Statistical Office of the European Communities (Eurostat)

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

1. Contact	
1.1 Contact organisation	Statistical Office of the European Communities (Eurostat)
1.2 Contact organisation unit	Unit C2: National accounts - production
1.5 Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update	
2.1 Metadata last certified	05 February 2009
2.2 Metadata last posted	
2.3 Metadata last update	05 February 2009

Reference metadata

3. Statistical presentation

3.1 Data description

National accounts are a coherent and consistent set of macroeconomic indicators, which provide an overall picture of the economic situation and are widely used for economic analysis and forecasting, policy design and policy making. Eurostat publishes annual and quarterly national accounts, annual and quarterly sector accounts as well as supply, use and input-output tables, which are each presented with associated metatdata.

Annual national accounts are compiled in accordance with the European System of Accounts - ESA 1995 (Council Regulation 2223/96). Annex B of the Regulation consists of a comprehensive list of the variables to be transmitted for Community purposes within specified time limits. This transmission programme has been updated by Regulation (EC) N° 1392/2007 of the European Parliament and of the Council. The domain consists of the following collections:

GDP and main aggregates. The data are recorded at current and constant prices and include the corresponding implicit price

1. Contact	8. Release policy	
2. Metadata update	9. Frequency of dissem	
3. Statistical presentation	10. Dissemination forn	
4. Unit of measure	11. Accessibility of doc	
5. Reference period	12. Quality management	19. Data revision
6. Institutional mandate	13. Relevance	20. Statistical processing
7. Confidentiality	14. Accuracy and reliability	21 Comment

Me

Metadata Standards

Dublin Core

The Dublin Core is an effective element set for describing a wide range of networked resources.

The Dublin Core standard comprises fifteen elements.

These are:

Title
Subject and keywords
Definition
Source
Language
Relationship
Scope
Creator / author
Published by
Contributors
Rights and management
History
Type
Format
Defined by

Meta data Studies in TURKSTAT

- We started to work by evaluating the current situation of TURKSTAT and inspecting the international literature and standards for metadata.

- Discussions for the best standard(s) to fit our needs were made.

- The studies of DDI for describing metadata for production have been initiated.

Meta dataStudies in TURKSTAT

- Metadata working group was set up in 2011

- Data and process standardization action plan was prepared

- Metadata Editor1.2 (DDI) was selected as the tool to catalogue the survey metadata in TurkStat.

Metadata Studies in TURKSTAT

- TURKSTAT metadata template was created by the metadata group

- The group modified the template, standardised the procedures about how the metadata is going to be entered to the template and code lists.

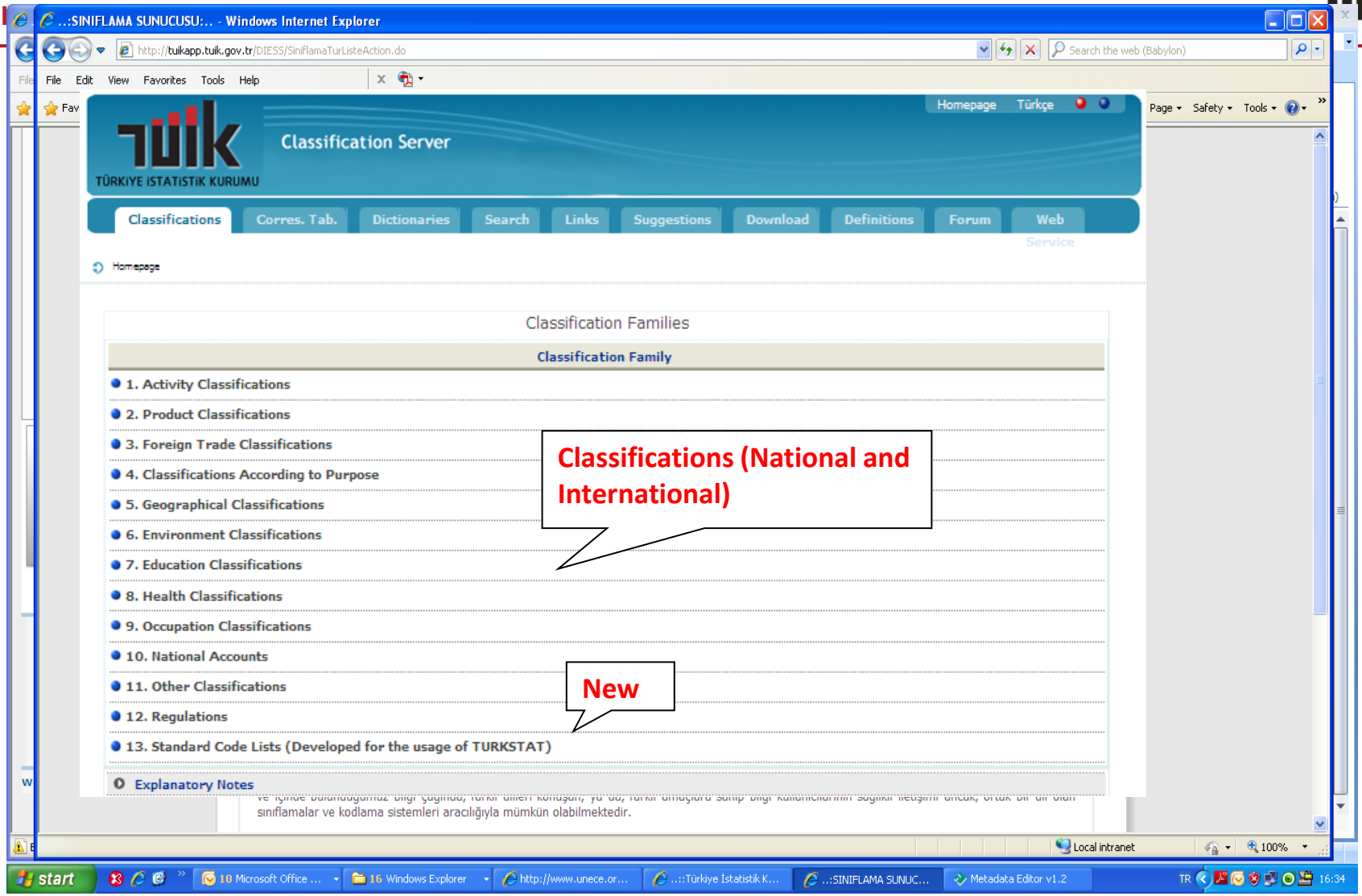
- Institution wide training was provided to all users

Metadata Studies in TURKSTAT

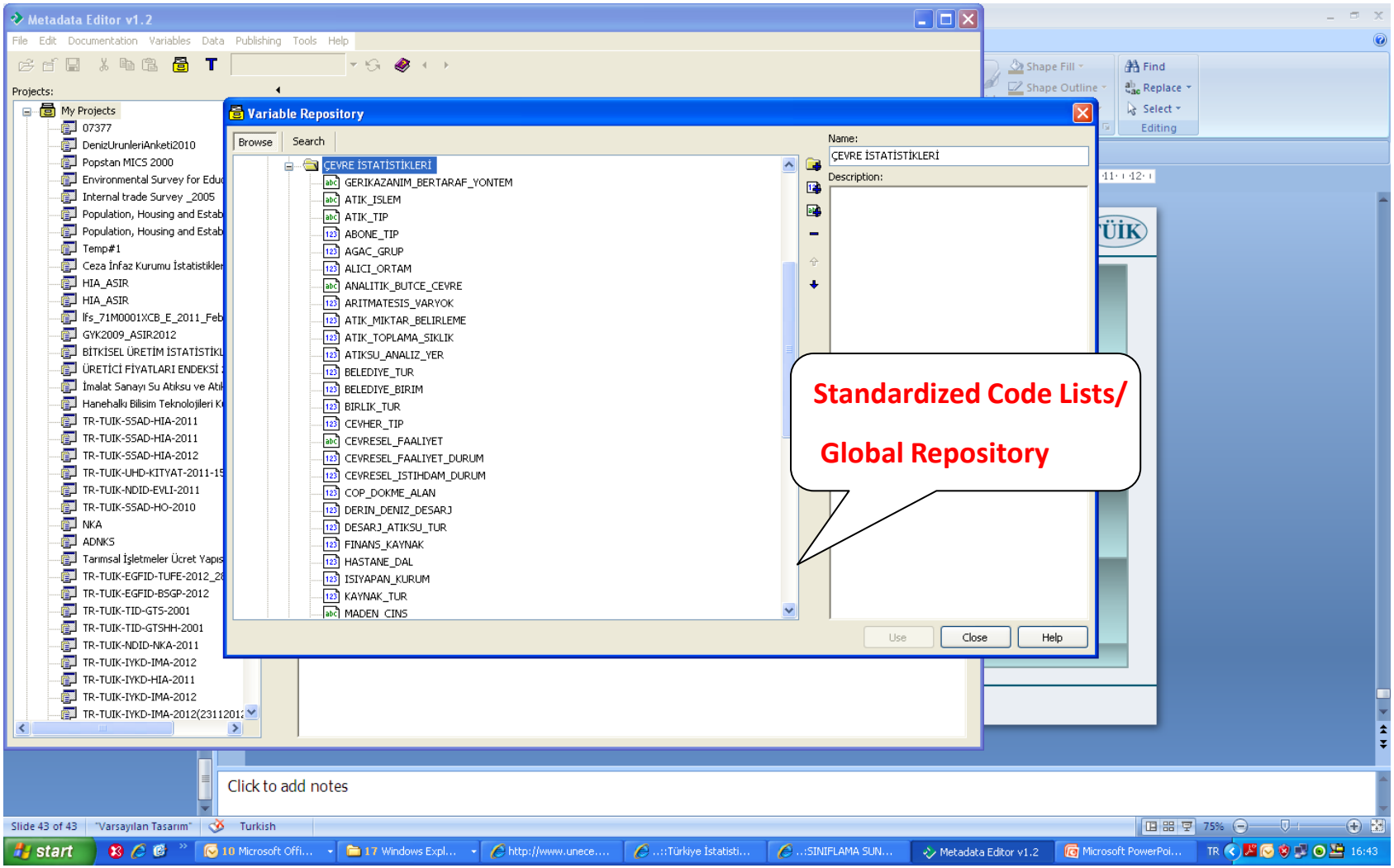
- This working group also extracted the code lists used in all business surveys, household surveys, agriculture and environment surveys, and tried to standardise the codes within each group of surveys as a first step. Then these code lists were brought together for all surveys.

- Standardization of a code lists were finalized for only the production databases.

- A catalogue of standard internal code lists was added to the web page for internal use only.



Classification Server, including a new section for the standard code lists



Metadata Editor _ Global Repository

Metadata Studies in TURKSTAT

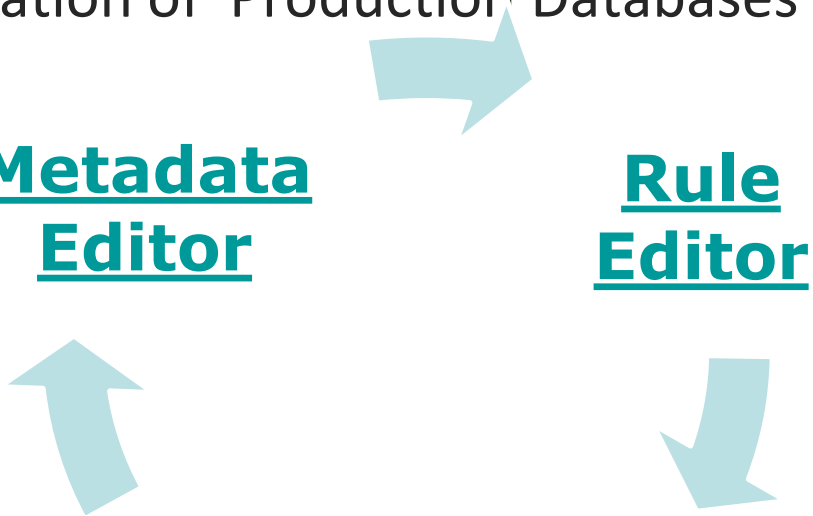
- The variable names were also listed and they are tried to be standardised according to the internally defined standard naming conventions. A catalogue of variable names was created for common variables.

- 12 pilot projects were entered in DDI format

- The metadata template was customized according to the pilot project result.
- The reference and structural metadata of current 125 surveys were collected DDI format and published on Intranet for only internal usage.

Metadata Studies in TURKSTAT

Standardization of Production Databases Process



Production Databases

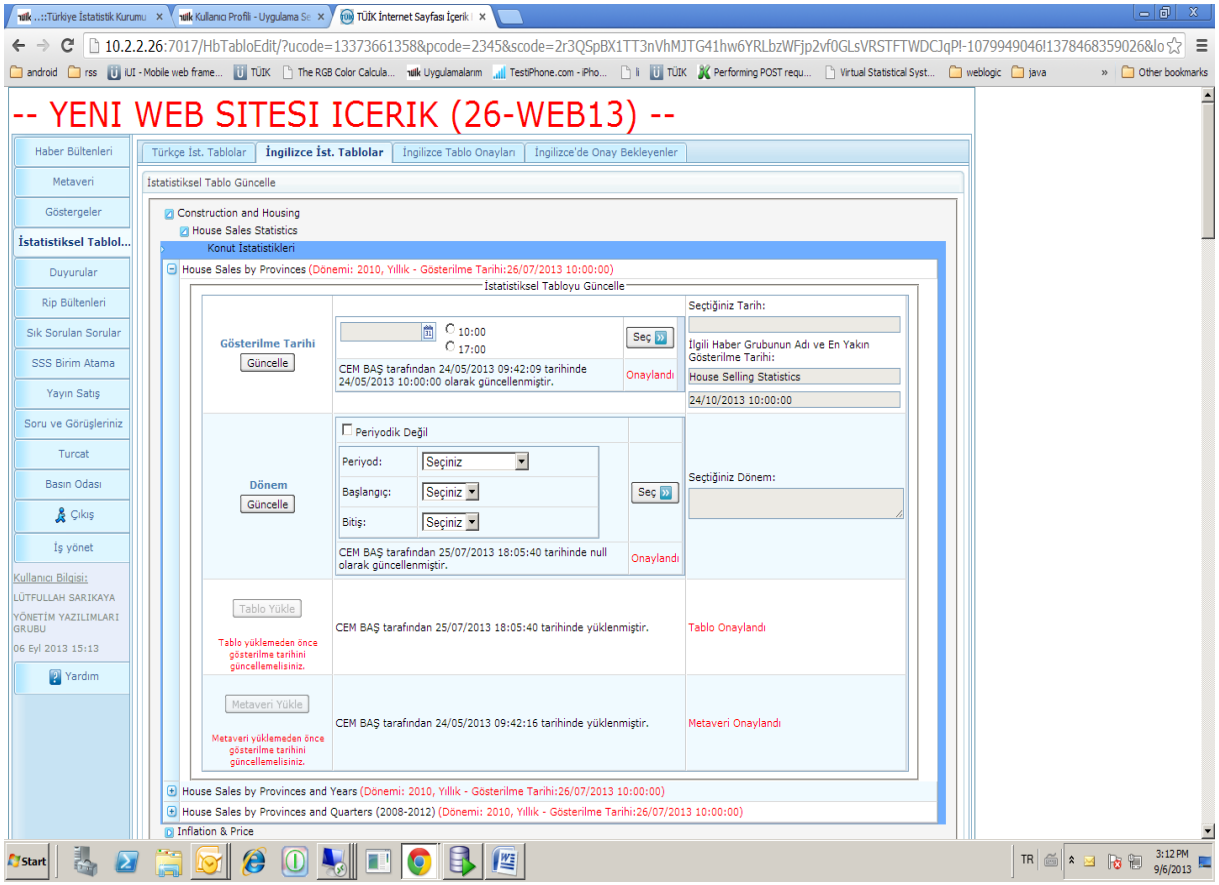
Current Situation

- Reference metadata about the products were published in the Turkstat's web page, mostly in SDDS format. Definitions of variables and concepts used in surveys were published in the web page for each product separately.

- Web Content Management Application is used for entering and updating metadata about the statistical theme and tables .

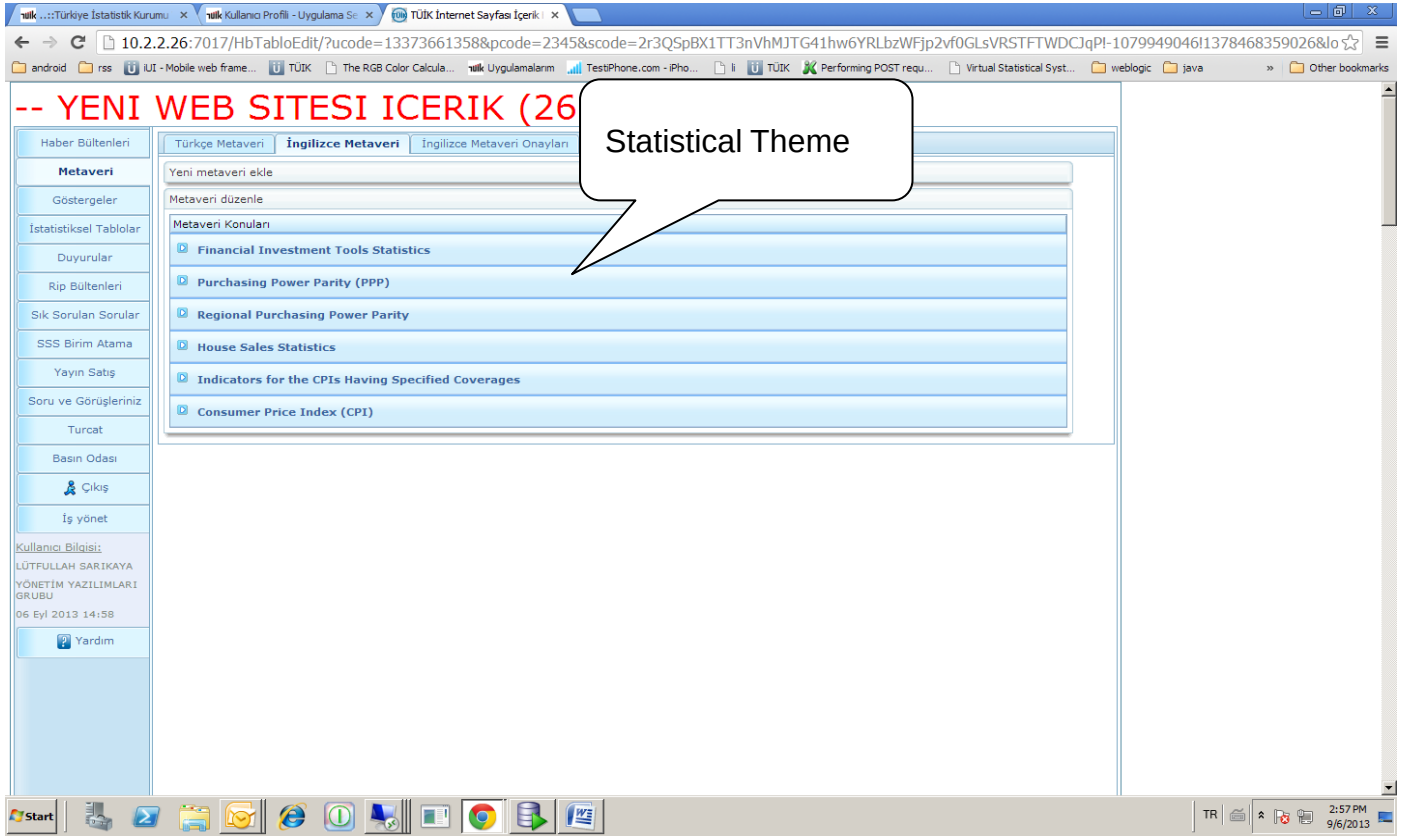
Current Situation

http://www.turkstat.gov.tr/



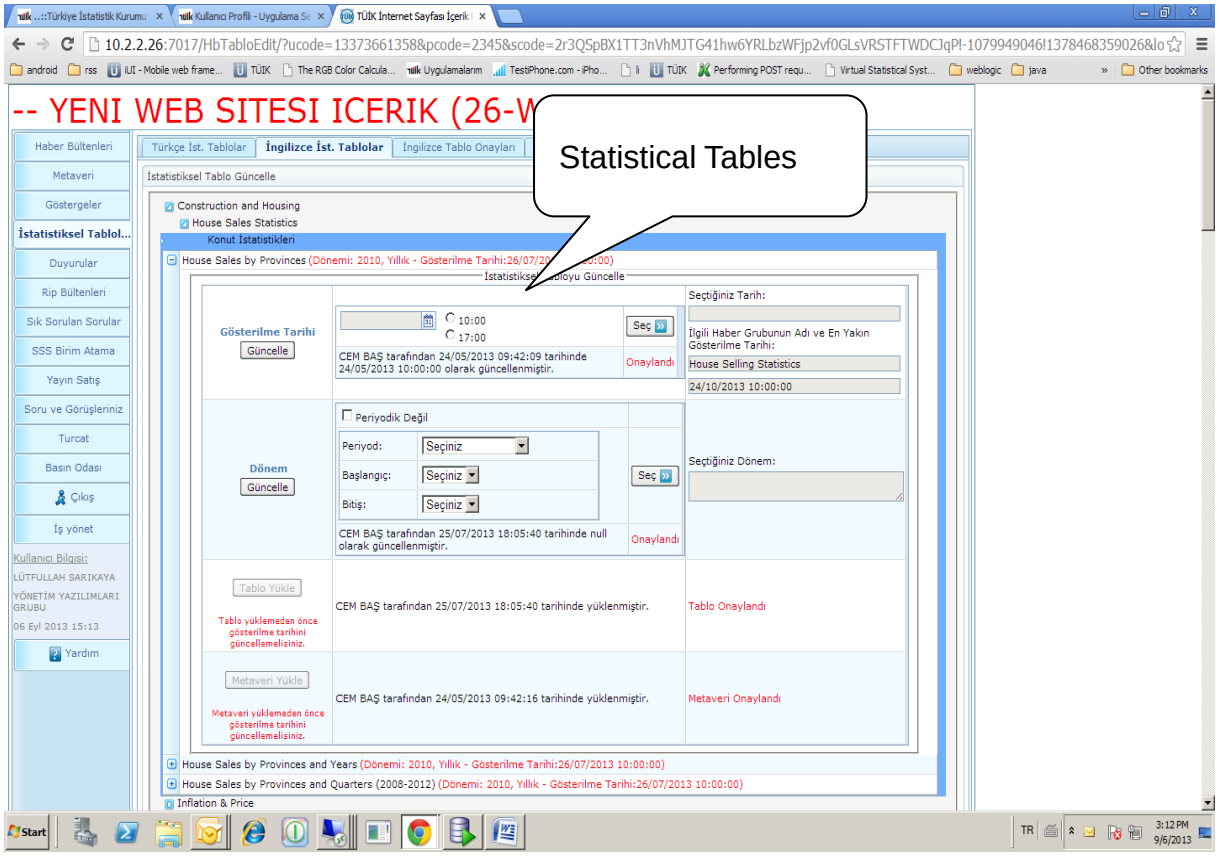
Interface of the Web Content Management Application

Current Situation



Interface of the Web Content Management Application

Current Situation



Interface of the Web Content Management Application

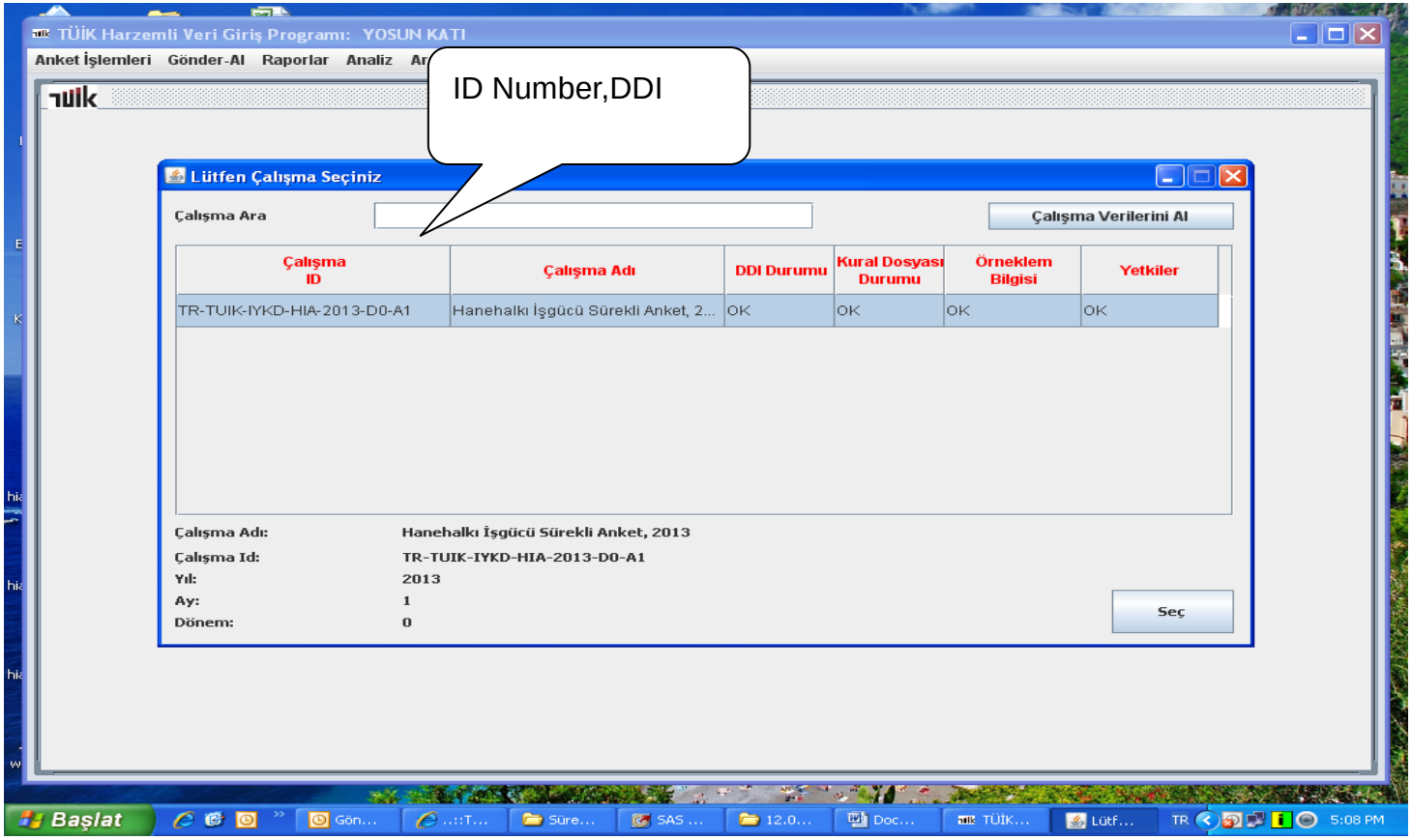
Current Situation

- There is a classification server which contains the national and international classifications.
- Establisment of central meta data system studies were started.
- By the current, reference and structural metadata information have been entered according to the standards in DDI format for the field works of surveys to be done in 2013.

Current Situation

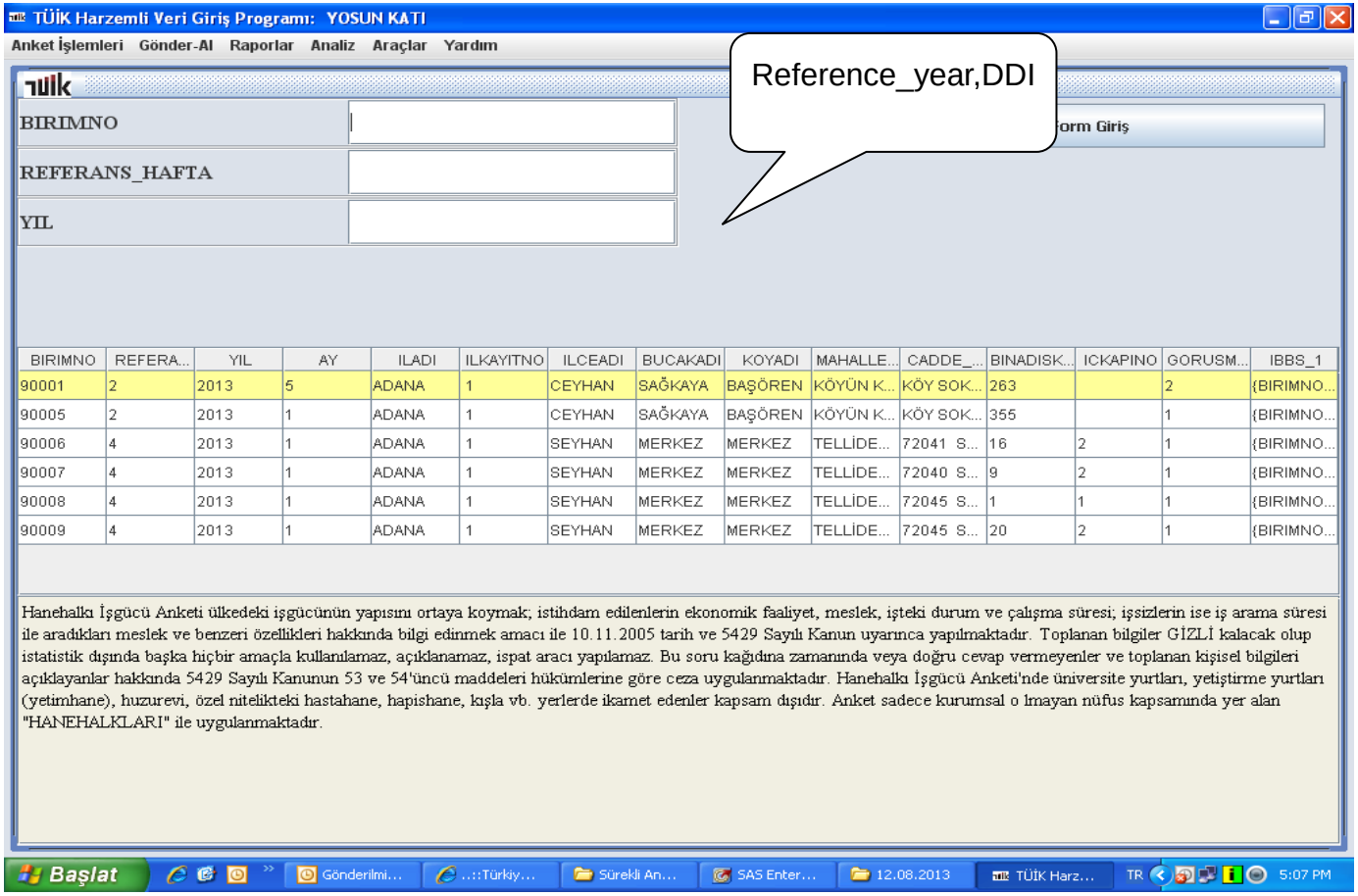
- Metadata-driven data entry applications are produced by a generic sofware.
- In this context, production databases have been standardized for 30 surveys and transferred to the generic software.
- Transfering of all researches to the metadata-driven production process is going on in coordination with subject matter divisions.
- Works of standardization about variables and Code lists in the Institutional databases are ongoing.

Current Situation



Interface of the generic data entry application

Current Situation



The interface of the generic data entry software

Experiences

- ❑ Metadata studies still stays as a difficult and important issue for NSI.
- ❑ A great deal needed on harmonisation of metadata in various processes and standards
- ❑ Developing a statistical metadata system is not just an issue for information technologies
- ❑ Commitment and continuous support of the top management
- ❑ Systematic and detailed up-to-date documentation

Key Points

- Think metadata for all the processes of statistics
- Use the international standards for **industrialisation**
- Do not think metadata as just the topic of IT
- Create a central metadata system, but with thinking **component oriented approach**
- Keep in mind the transformation and integration of any kind of metadata Metadata studies still stays as a difficult and important issue for NSI.

Future Plans for Development

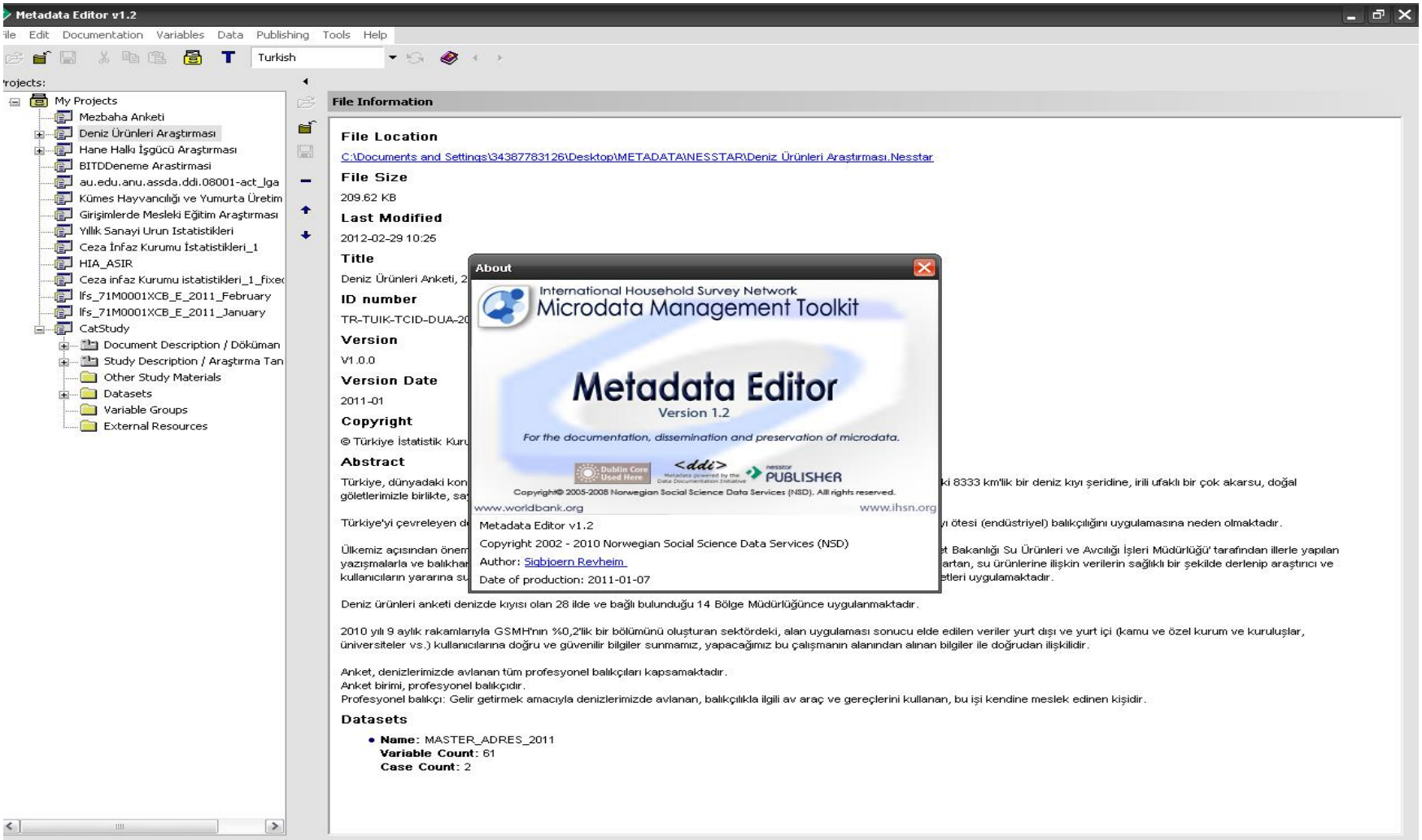
- To prepare metadata strategy document
- To develop central and integrated meta data repository including all statistical production system.
- To capture metadata in all processes of statistical production at the time it occurs.
- To promote the documentation in all processes of statistical production.
- To develop variable and concept repository
- To review processes and optimize the process workflows

Metadata Editor

- Metadata Editor is a software used to create the metadata according to international metadata standards (DDI and Dublin Core).

- Metadata Editor ensures to enter both the reference and the structural metadata and also allows to keep entered metadata as XML files in DDI format.

Metadata Editor



Metadata Editor

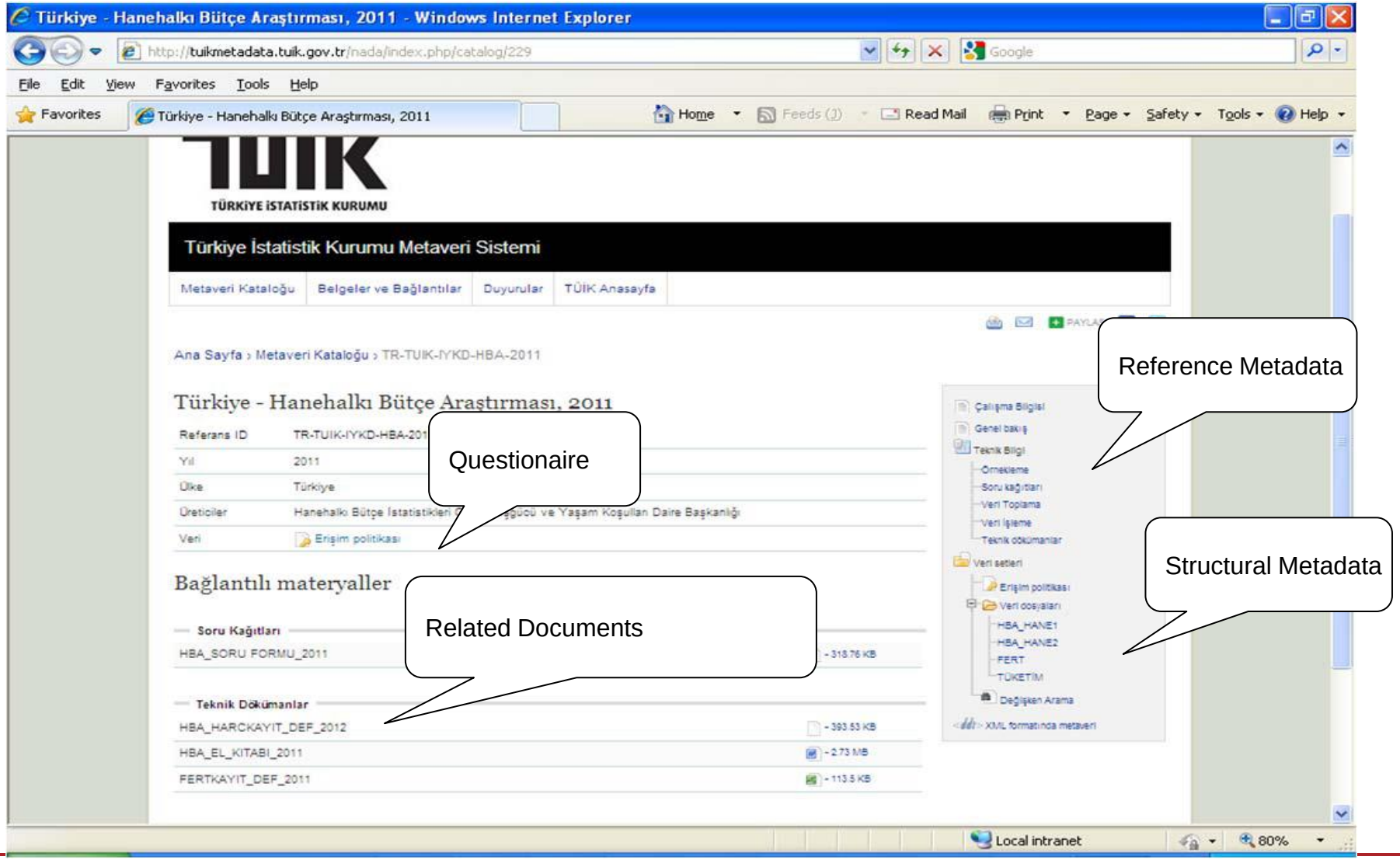
- XML file

Metadata Publishing Platform:NADA

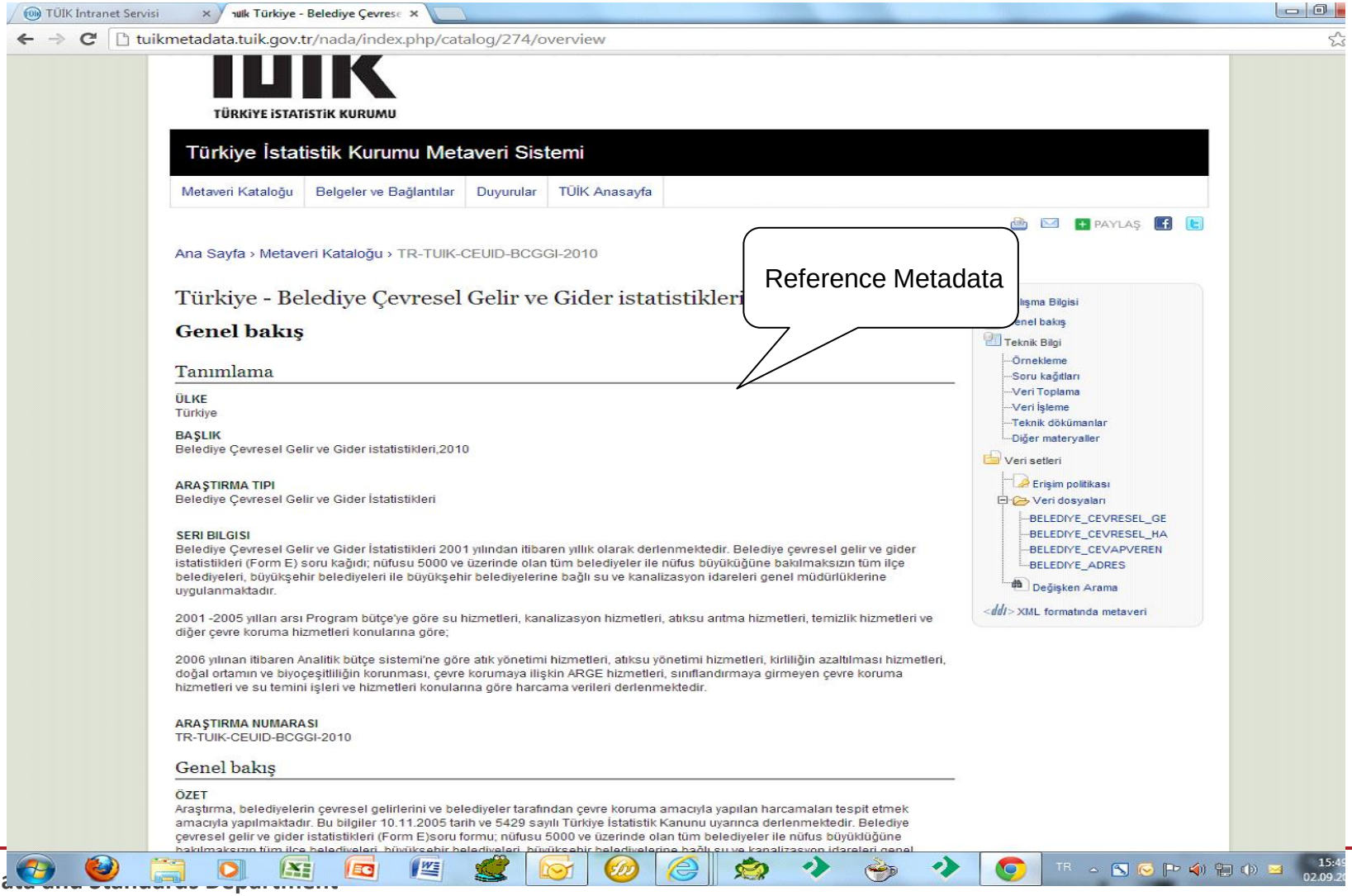
- NADA is the metadata publishing platform to browse, search, compare, apply for access, and download relevant census or survey information.
- It is designed to provide metadata information via web.
- NADA uses the Data Documentation Initiative (DDI), XML-based international metadata standard.

Metadata Publishing Platform:NADA

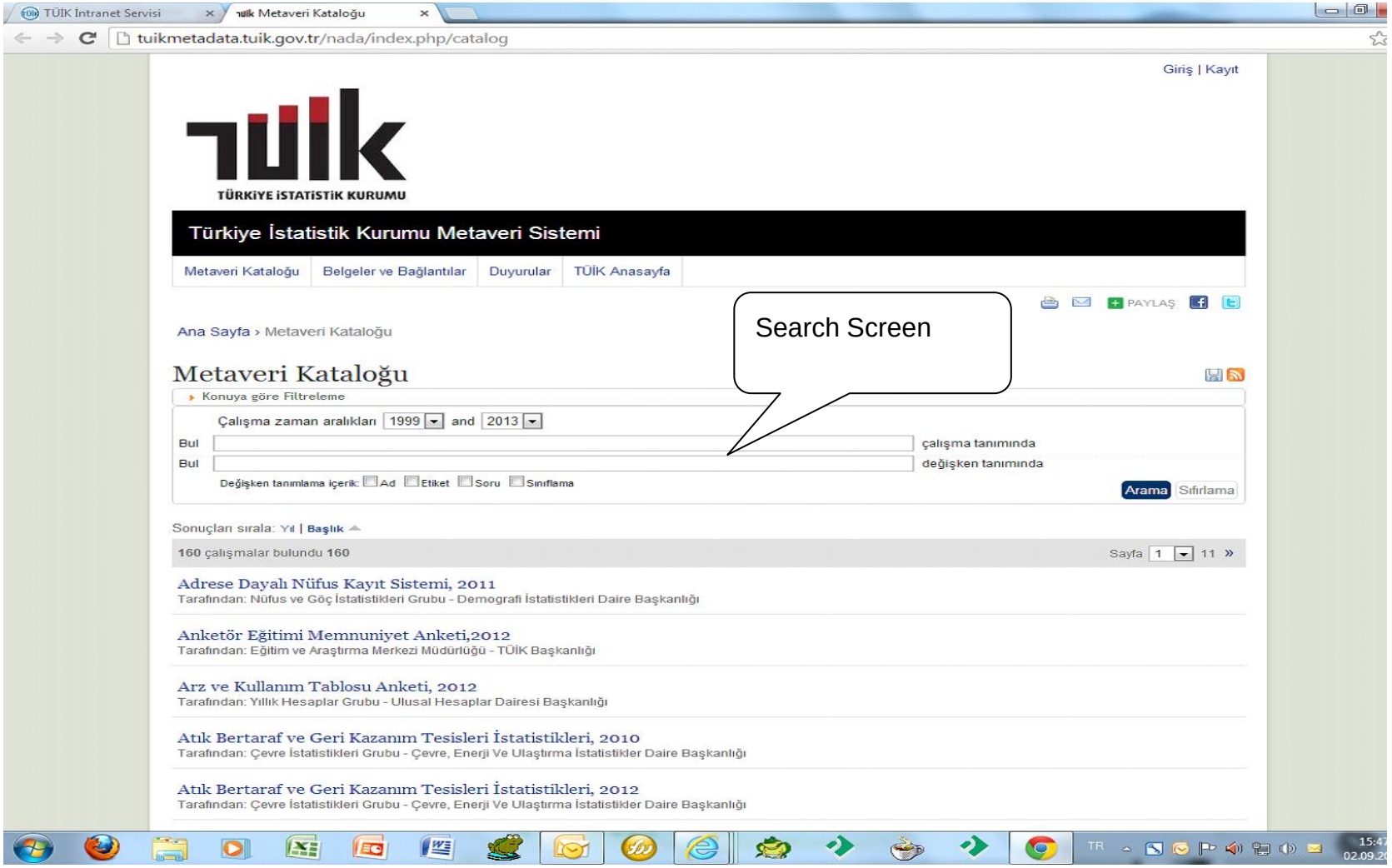
Meta data Publishing Platform:NADA



Meta data Publishing Platform:NADA



Metadata Publishing Platform:NADA



References

[1] http://www.ddialliance.org/

[2] http://www.sdmx.org

[3] Arofan Gregory, Open Data Foundation, 2011, The Data Documentation Initiative (DDI): An Introduction for National Statistical Institutes

[4] Arofan Gregory, Pascal Heus, Open Data Foundation, 2007, DDI and SDMX:

Complementary, Not Competing, Standards

[5] Wendy Thomas, Minnesota Population Center, 2010, Course on DDI 3: Putting DDI to Work for You

References

[6] Gökhan, GÜRGEÇ, High Level Seminar for Eastern Europe , Caucasus and Central Asia Countries (ECCA) on 'Qality in Statistics Metadata' ,Metadata Studies in TURKSTAT, 2013

[8] Arofan Gregory , METIS, 2011, Overview of DDI

[9] Mary Vardigan, Fourth Conference of the European Survey Research Association, 2007,Overview of the Data Documentation Initiative (DDI) Metadata Standard: Goals and Benefits

[10] Özkan,Deniz ,Work Session on Statistical Metadata, Implementation of GSBPM and Metadata Studies in Turkish Statistical Institute,2013

Thank you very much for your attention!