Central Dissemination Project

1.4.2014
(Muscat, Oman)
OUTLINE

• Introduction
• Current System
• Deficiency Of Current System
• Advantages Of Central Database
• Indicators Model For Central Dissemination
• Data Process Of Central Dissemination Database
• Tables of Project
• Web Application for Independence Statistical Indicators
• GVYU Application and MEDAS Application
INTRODUCTION

- Dissemination databases are the summary aggregated data repository which are created from institutional databases according to the information request.

- Aggregated time series data created from TurkStat institutional databases.

- They are accessible from Web Page.

- Dissemination databases are for end users but not for beginners.

- Users take data from database according to the given criteria. Reports are dynamically produced.
• While disseminating data on the web page, data **confidentiality** are considered. All data at the institutional database are not included at dissemination database

• TR, Nuts1 (12 region), Nuts2 (26 region), Nuts3 (81 city),

• Town and Municipality level data is accessible through Institutional Database
CURRENT SITUATION

- A java application for selecting report parameters
- Sending parameters via URL to report server
- Oracle Reports Builder handles the requests
A Current Java Application

![Livestock Statistics](http://www.tuik.gov.tr/hayvancilikapp/hayvancilik_jing.zul)

**Livestock Statistics**

- **Number of animals and animal products**
- **Number of animals according to age groups**
- **Number of animals slaughtered, meat and hide**

**Year Selection**

- 2010
- 2009
- 2008
- 2007
- 2006
- 2005
- 2004

**Table Selection**

- Bovine Animals
- Sheep and Goats
- Poultry
- Apiculture
- Sericulture
- Equines
- Pig
- Camel

**Table format:**

- [Excel](http://rapor.tuik.gov.tr/reports/rwservlet?hayvancilik)
- [PDF](http://rapor.tuik.gov.tr/reports/rwservlet?hayvancilik)

**Information and remarks:**

Note: Data on red meat, chicken meat and egg production have been published on the website through monthly news releases starting from January 2010.

- [Clean Screen](#)
- [Make Report](#)
... **Why** do we need to design a central database for dissemination?
Deficiency of Current System:

- For each application gathering the same data with different ways
- Different database designs
- Different application for each subject of statistics
Advantages of Central Database:

- Reduce actual reporting burden
- Administrative simplification
- Generic data
- Uniform report helps the web service development and administration in the future.
Indicators Model for Central Dissemination

• **Indicator**: Identifier information except time (year, month, quarter) and location (region).

• We assume that Each given statistic is an indicator.

*Example:*

*Exchange ratio of unemployment rate for male population in Ankara on March, 2000, is 0.3.*
Indicators Model (Cont’d)

Data is composed of:

Time – Level – **indicator** - Value – Metadata

Examples of Population indicators

Population
Male population
Male - literate – population
Male - illiterate – population
Female - literate – population
Indicators Model (Cont’d)

Example:

For unemployment suppose that having 3 dimensions.

Gender : 2 codes,
Education : 10 codes,
Occupation : 300 codes

Indicators:

unemployment with no dimension : 1
With 1 dimension : $2 + 10 + 300 = 312$
With 2 dimensions : $2 \times 10 + 2 \times 300 + 300 \times 10 = 3620$
With 3 dimensions : $2 \times 10 \times 300 = 6000$
Indicators Model (Cont’d)

- Each indicator should have identification code.

Indicator : Measurement + Dimensions

- There are also indicators that don’t include any dimension (independence indicator)

Measurements and Dimensions **MUST** be coded.

A classification server should be used for all projects for dimension codes
Data Process of Central Dissemination Database

Statistical units prepare the data in a common format that is suitable for the system.

Prepared data is transferred to the institutional central database via indicator Data Management app.

Data control, approval process

Institutional Central Database

Central Dissemination Database
There is a standard fact table structure for each project

Dimension codes should exist in the **classification server**

Updating of the Classification server should affect all projects So be Careful!
Fact Table Columns

1. **YIL, AY, DONEM**: Time
2. **DUZEY, DUZEY_KOD**: Place
3. **FACT_OLCUM_NO**: measurement id
4. **GOSTERGE_NO**: Indicator id
5. **DEGER, DEGER_AD**: Value
6. **GENEL_ALAN**: the reason of new data (data / metadata)
7. **VERI_TANIM**: flag for hiding data
8. **AKTIF_PASIF**: active/passive
9. **DIMENSION ALANLAR**: dimension columns
10. **ACIKLAMAn, ACIKLAMAn_ING** (n:max 10): Metadata
METADATA

- For each indicator (in indicator table)
- For each record
- There are 10 metadata column in fact table
- METAVERI_BILGI table contains metadata
TABLES OF PROJECT

[Diagram of database schema showing tables and relationships]
Classification server’s Table structure

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATIR_ID</td>
<td></td>
</tr>
<tr>
<td>SURUM_ID</td>
<td></td>
</tr>
<tr>
<td>SURUM_ADI</td>
<td></td>
</tr>
<tr>
<td>SURUM_ADI_EN</td>
<td></td>
</tr>
<tr>
<td>KOD</td>
<td></td>
</tr>
<tr>
<td>TANIM</td>
<td></td>
</tr>
<tr>
<td>TANIM_EN</td>
<td></td>
</tr>
<tr>
<td>UST_KOD</td>
<td></td>
</tr>
</tbody>
</table>
Generated data, Waiting until release time

Release time

After release time

The data displayed on the web page

indicator’s Identification code
APPLICATIONS

There are 3 applications

1 - Web application for Independence Statistical Indicators
   (now includes 66 important indicators)

2 – GVYU - Data Management application for producing and managing data

3- MEDAS (Central Dissemination Application) : For querying
   the central dissemination database
1- Web Application for Independence Statistical Indicators
1- Web Application for Independence Statistical Indicators (Cont’d)
2- GVYU Application

• This application is used by TurkStat personels

• This is Central Data Management Application

• Java is used, web based application

Missions:

• defining indicators and indicator’s metadata, generating indicator ID

• Controlling department’s data before deploying

• Deploying data into temporary tables

• Approval mechanism

• Deploying data into live dissemination table (Preparing MEDAS data)
2- GVYU Application (Cont’d)
2- GVYU Application (Cont’d)
3- MEDAS APPLICATION
(Central Dissemination Application)

Subject: Foreign Trade Indexes

Measurement:
Base year: 2010

Choose a Measurement:
- Export unit value index (2010=100)
- Import unit value index (2010=100)
- Export volume index (2010=100)
- Import volume index (2010=100)

Number of chosen Indicators: 0
Number of chosen Regions: 0
Number of chosen Time: 0
3- MEDAS APPLICATION (Cont’d)
3- MEDAS APPLICATION (Cont’d)
3- MEDAS APPLICATION (Cont’d)
What did we do until now in this project?

- 66 independent indicators are selected and are accessible through TurkStat websites (map and graphic is available for regional data)
- For now, the data is given up to province level. (Nuts3) (Also TR, Nuts1, Nuts2, Town and Municipality level data is available in Institutional DB)
- The Central Dissemination makes the comparison between different statistical subjects’ possible.
- Till now 5 statistical units’ data is uploaded for the central dissemination and could be queried by MEDAS Application (Not accessible on web page yet)
SUMMARY

✓ We use Indicator Model in Central Dissemination Database

✓ Indicators are generated using dimensions using a software program

✓ Central Dissemination System reduces reporting burden

✓ Central Dissemination System simplify administration of data

✓ End users can get dynamic reports, pivot reports and can export report to Excel

✓ We are developing the system, not finished yet
Important Notes

Separating the databases is important and necessity. **Production** database should include current data. **Institutional** database should contain both current and historical data. **Dissemination** databases should include only disseminated data.

The tables structures at the Institutional DB will be similar to Production database. For generating **aggregated data** at Institutional database, Matealized Views may be created or SQLs are used to generate new tables.

**Standardization** of codes is essential

**Metadata** system should be considered

**Central dissemination is mandatory**
Thank you & Questions ?