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 takes data from database according to the given criterias.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

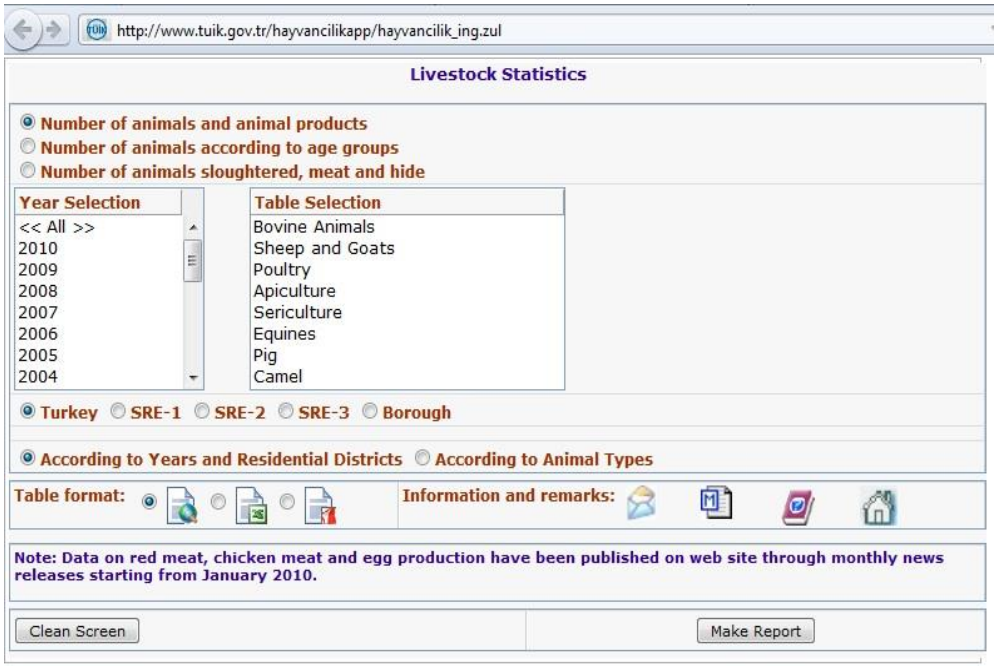
- TR, Nuts1 (12 region), Nuts2 (26 region), Nuts3(81 city),

- 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



http://rapor.tuik.gov.tr/reports/rwservlet?hayvancilik...

… **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 –populatin
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*10 + 2*300 + 300*10 = 3620
With 3 dimensions : 2*10*300 = 6000

Indicators Model (Cont'd)

- Each indicator should have identification code.

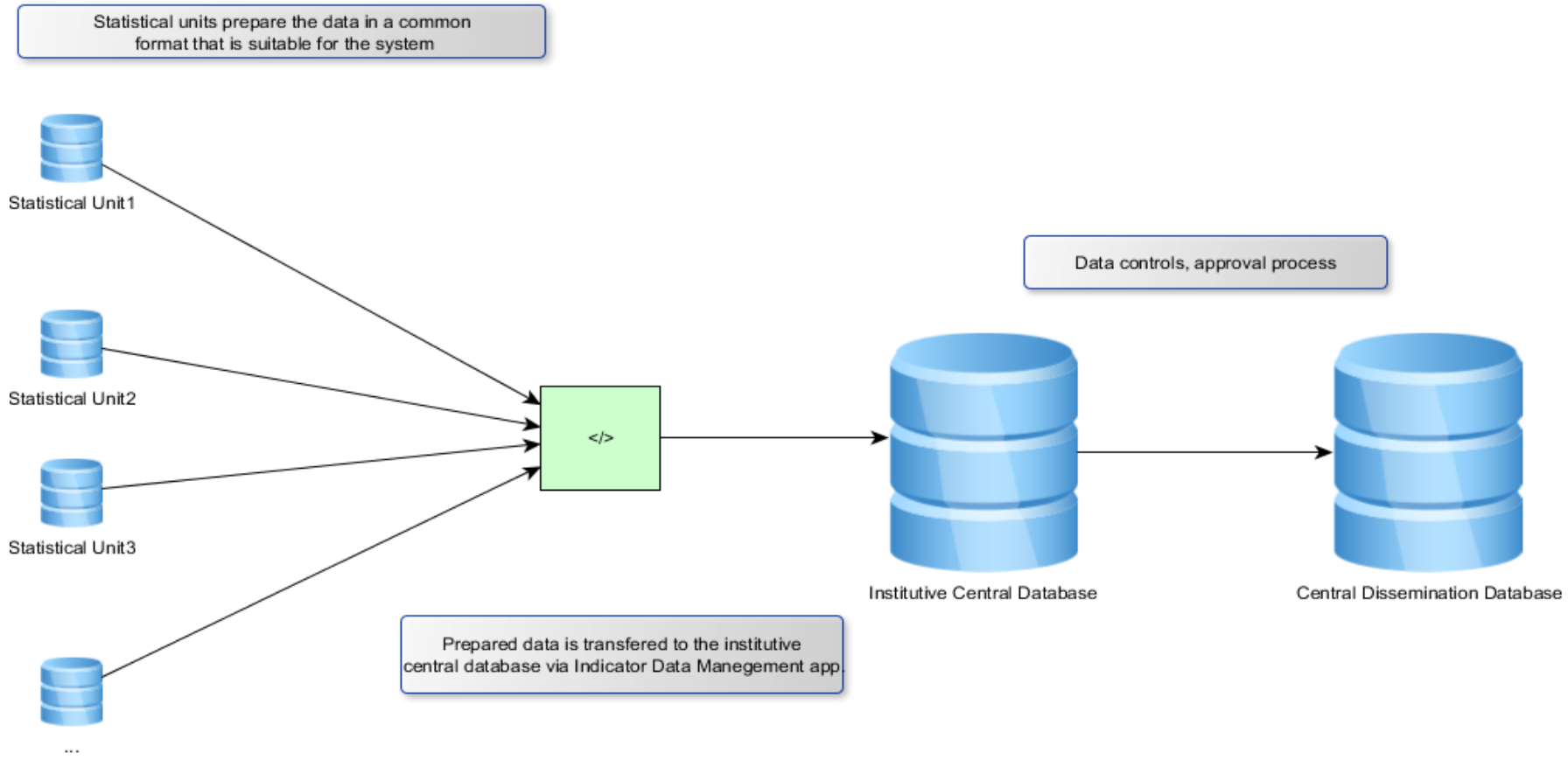
Indicator : Measurement + Dimensions

- 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



Fact Table Structure

time

place

value

dimensions

YIL	AY	DON...	DUZEY	DUZEY_KOD	FACT_OLCUM...	GOSTERGE_NO	DEGER	DEGER_AD	GENEL_ALAN	VERI_TANIM	AKTIF_...	EGITIM	CINS...	KIR_KE...	ACI
2010	0	0	3	42	1	-999	243088	-999	-999	-999	0	-999	1	-999	260
2011	0	0	2	TR71	1	-999	248154	-999	-999	-999	0	-999	1	-999	
2011	0	0	2	TRB1	1	-999	249749	-999	-999	-999	0	-999	2	1	
2010	0	0	2	TRB1	1	-999	253520	-999	-999	-999	0	2	2	1	
2010	0	0	2	TRA2	-999	URF-GK1244-O42	254729	-999	-999	-999	0	-999	-999	-999	
2010	0	0	3	7	-999	URF-GK1244-O42	255644	-999	-999	-999	0	-999	-999	-999	

There is a standart 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 !

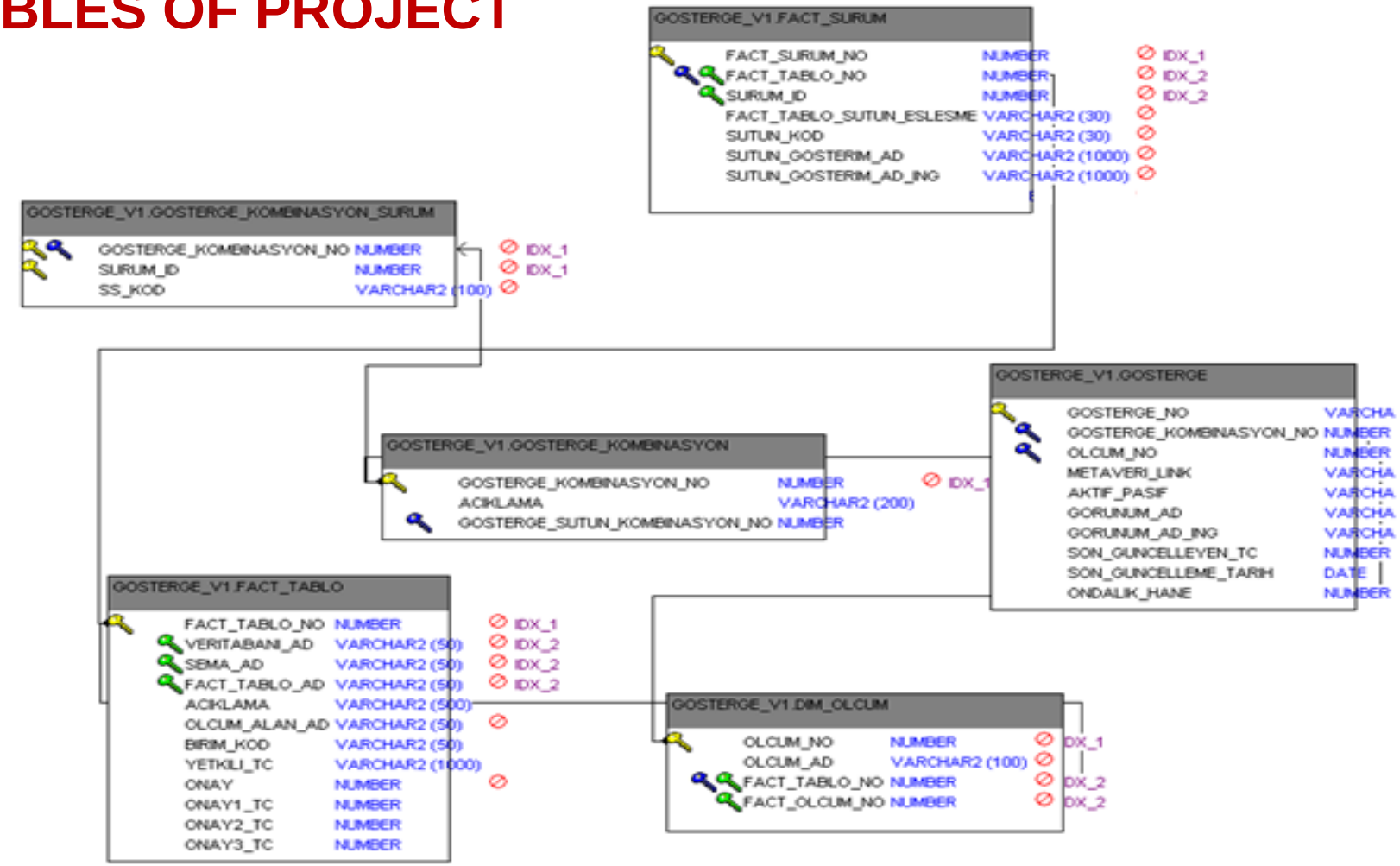
Fact Table Columns

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



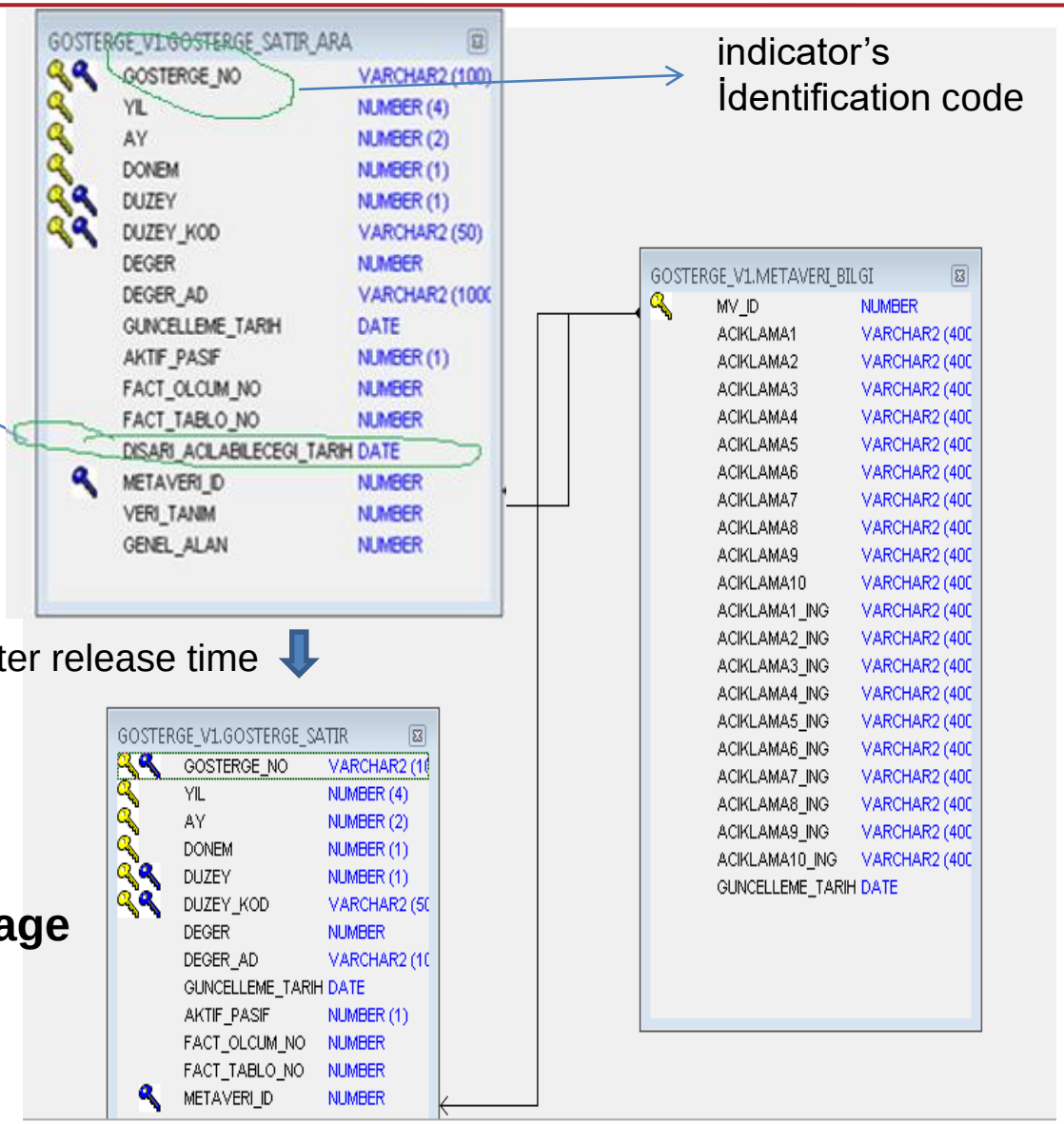
Classification server's Table structure

Column Name
SATIR_ID
SURUM_ID
SURUM_ADI
SURUM_ADI_EN
KOD
TANIM
TANIM_EN
UST_KOD

Generated data,
Waiting until release time

indicator's
İdentification code

Release time



The data
displayed on the web page

APPLICATIONS

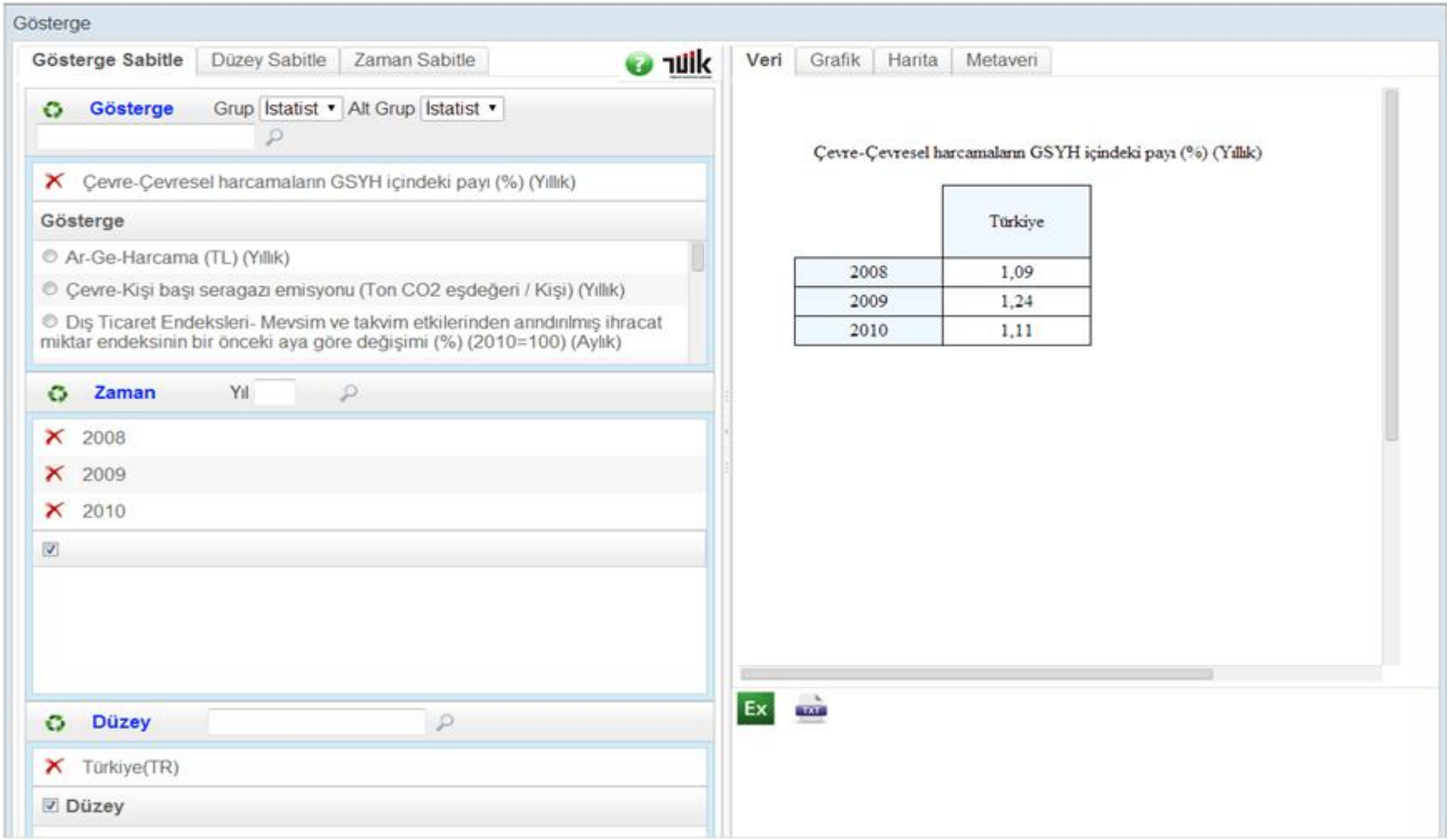
There are 3 applications

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

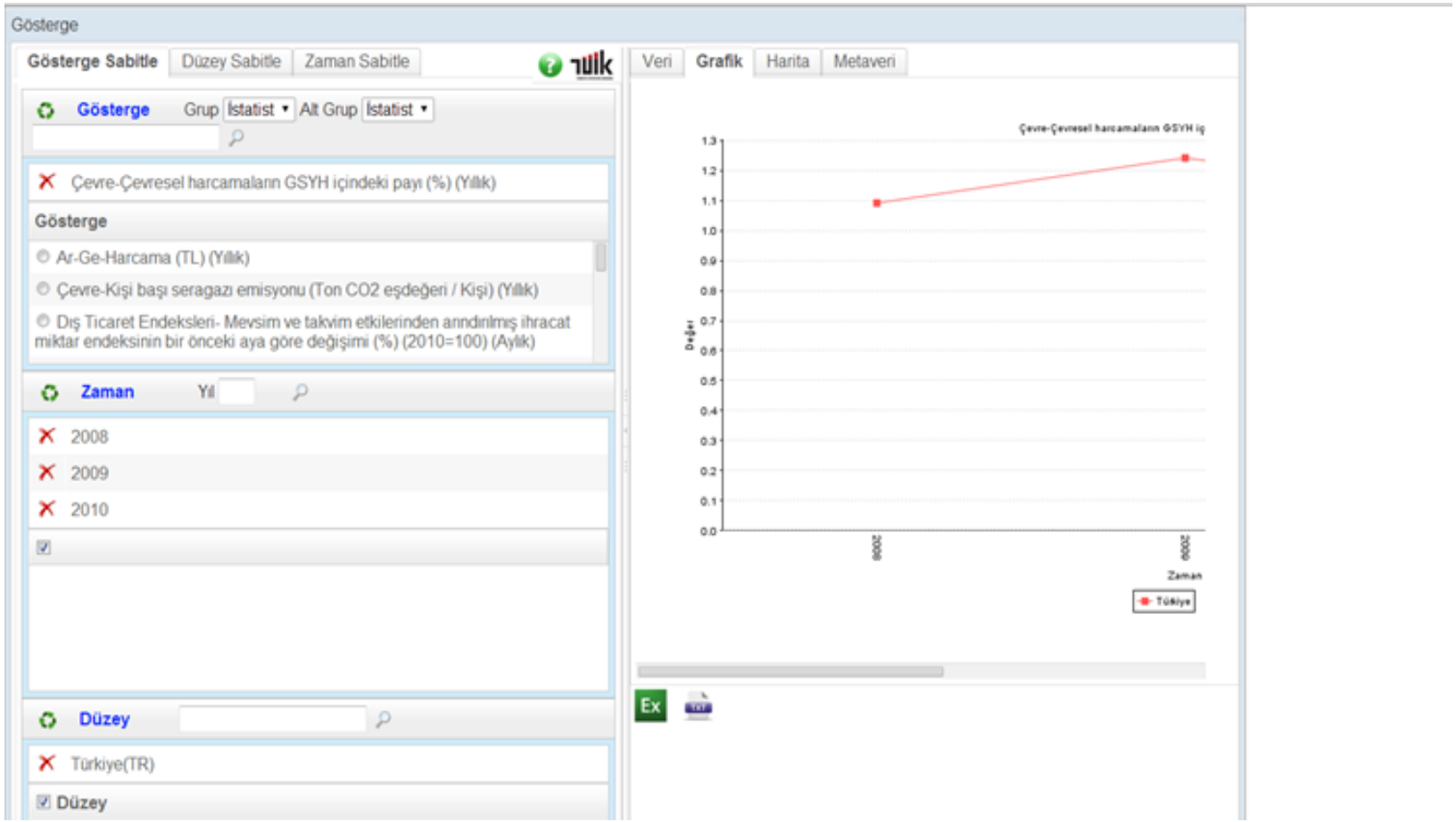
2 – GVYU - Data Management application for producing and managing data

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

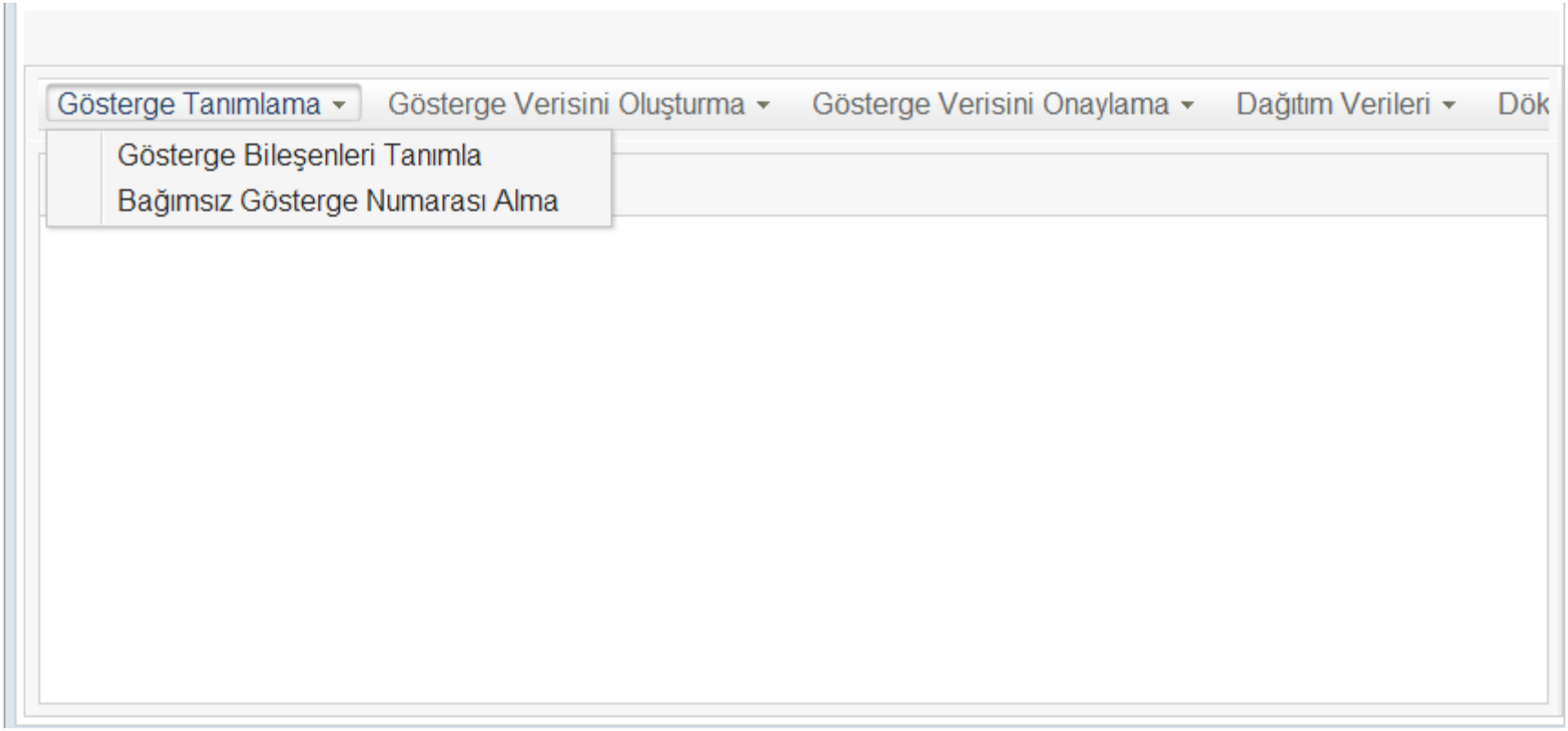
•Deploying data into temporary tables

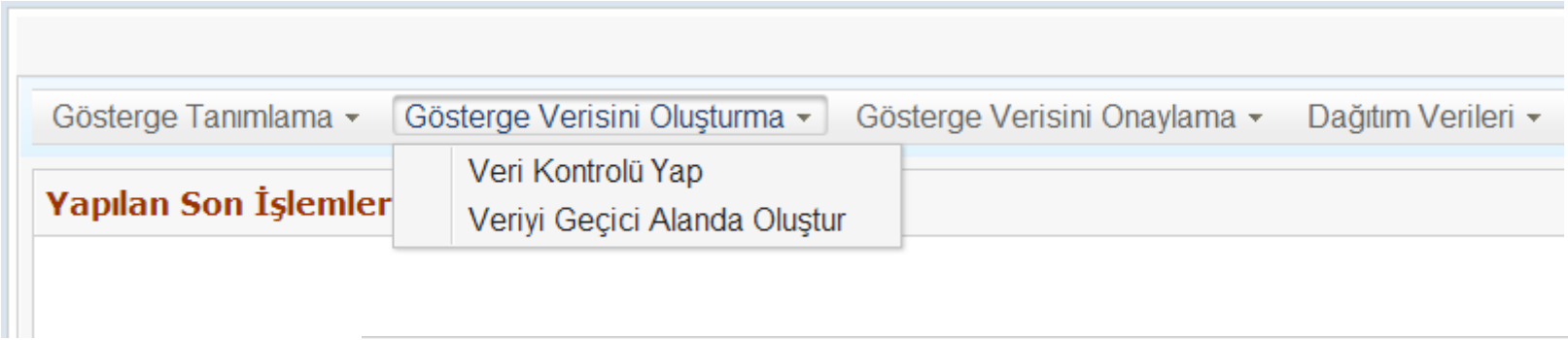
•Approval mechanism

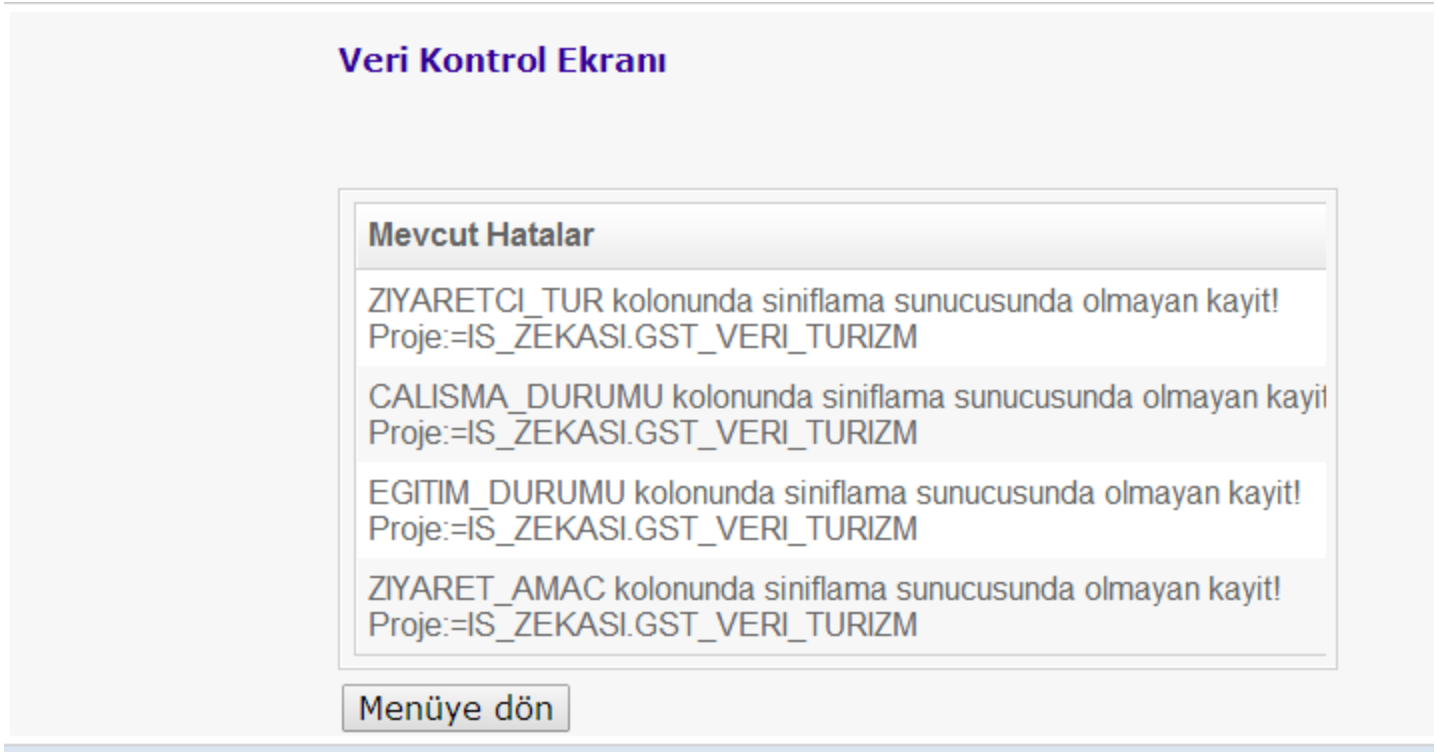
•Deploying data into live dissemination table (Preparing MEDAS data)

2- GVYU Application (Cont'd)

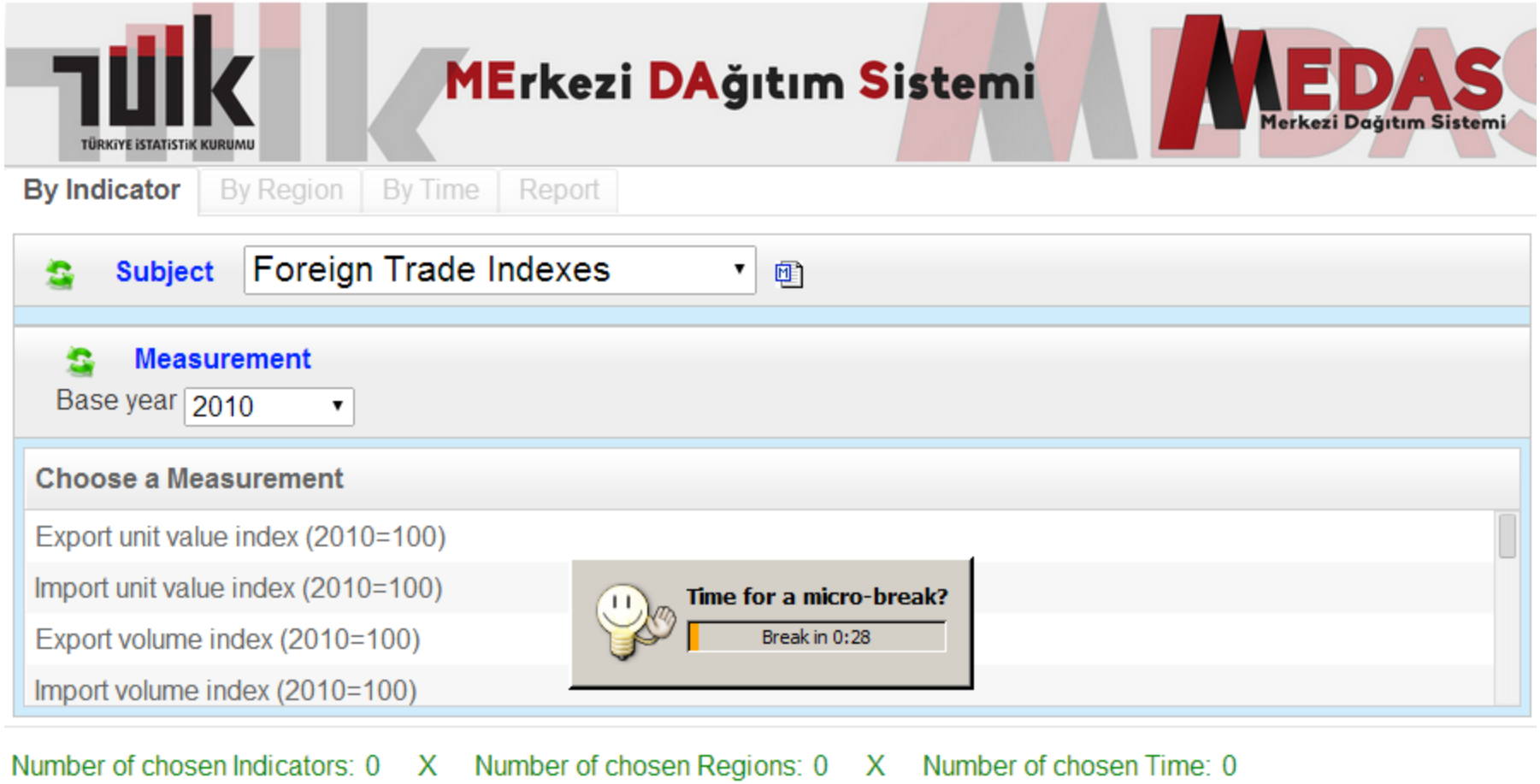
2- GVYU Application (Cont'd)




Gösterge Tanımlama
Gösterge Verisini Oluşturma
Gösterge Verisini Onaylama
Dağıtım Verileri
Veri Kontrolü Yap
Veriyi Geçici Alanda Oluştur
Yapılan Son İşlemler

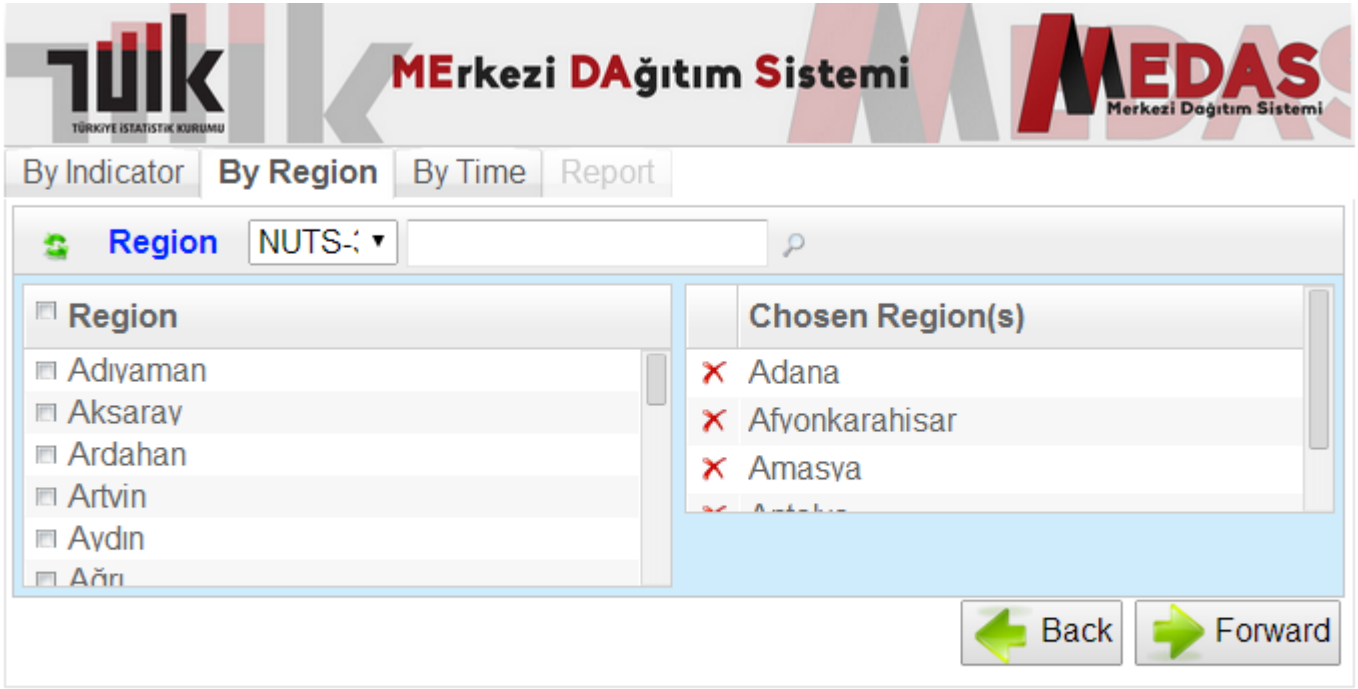

Veri Kontrol Ekranı
Mevcut Hatalar
ZIYARETCI_TUR kolonunda siniflama sunucusunda olmayan kayit!
Proje:=IS_ZEKASI.GST_VERI_TURIZM
CALISMA_DURUMU kolonunda siniflama sunucusunda olmayan kayit
Proje:=IS_ZEKASI.GST_VERI_TURIZM
EGITIM_DURUMU kolonunda siniflama sunucusunda olmayan kayit!
Proje:=IS_ZEKASI.GST_VERI_TURIZM
ZIYARET_AMAC kolonunda siniflama sunucusunda olmayan kayit!
Proje:=IS_ZEKASI.GST_VERI_TURIZM
Menüye dön

3- MEDAS APPLICATION (Central Dissemination Application)



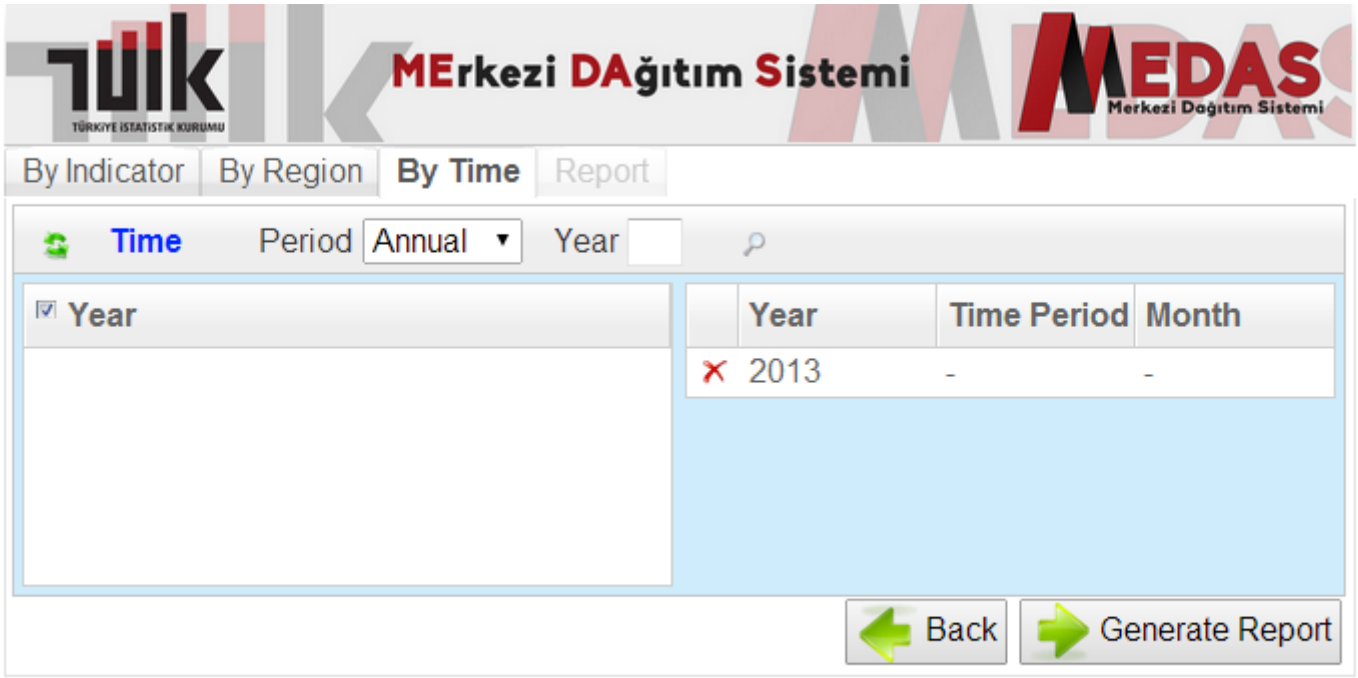
3- MEDAS APPLICATION (Cont'd)

3- MEDAS APPLICATION (Cont'd)



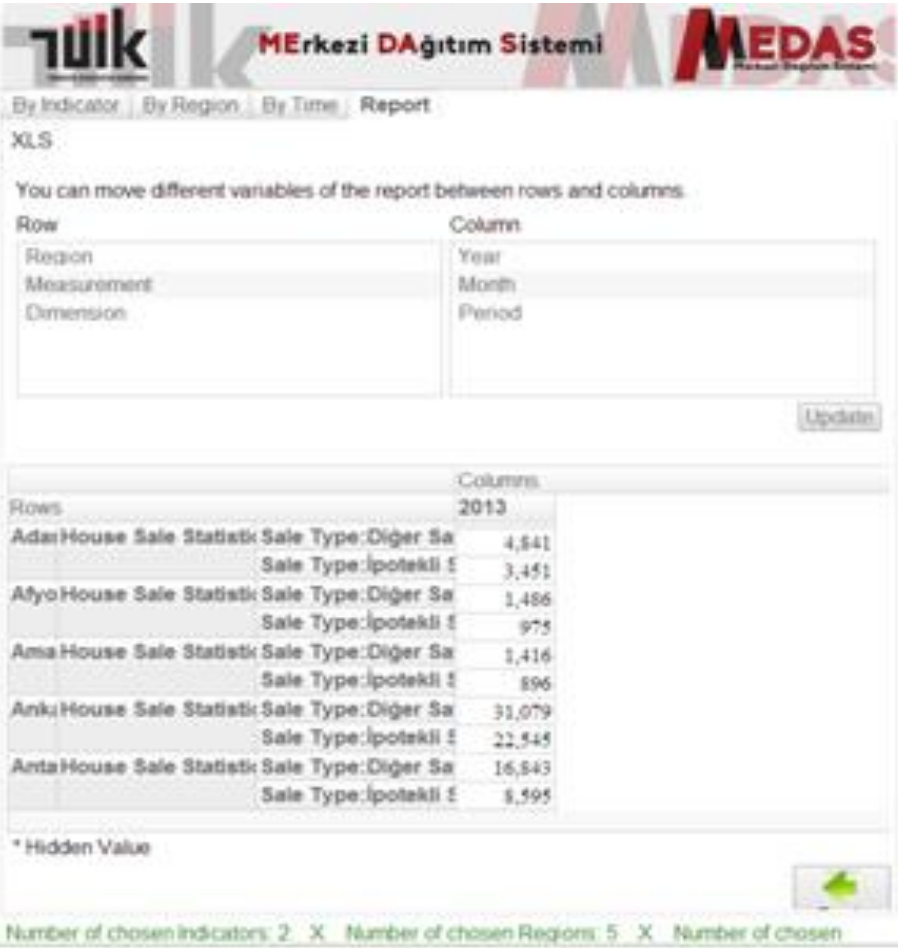
Number of chosen Indicators: 2 X Number of chosen Regions: 5 X Number of chosen Time: 1

3- MEDAS APPLICATION (Cont'd)



Number of chosen Indicators: 2 X Number of chosen Regions: 5 X Number of chosen Time: 1

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 Instutional DB)
- The Central Dissemination makes the comparision 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

Seperating 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 SQL s are used to generate new tables.

Standardization of codes is essential

Metadata system shoud be considered

Central dissemination is mandatory

Thank you & Questions ?