A corporate approach to processing microdata in Eurostat

Pál JANCSÓK and Christine WIRTZ
Eurostat Unit B4
Agenda

- Introduction
- Generic SAS Tool (GSAST) architecture
  - Microdata processing
  - Architecture
- Metadata in GSAST
  - Structure
  - Management
- Possible extension of GSAST to data providers
  - Current processing
  - Self service
- Conclusions
Introduction

- Eurostat is the statistical office of the European Union and a General Directorate of the European Commission.

- Eurostat is the central institution of the European Statistical System (ESS) - a network of National Statistical Institutes (NSI) from all EU and EFTA Countries.

- Eurostat's mission: to be the leading provider of high quality statistics on Europe.
Introduction

Streamline the statistical production:

**Internal efforts**

- Harmonisation, and consolidation.
- Modern, generic tools for data processing.

**External efforts**

- Cooperation with data providers, sharing infrastructure.
- Remote access for data quality.

**Internal G**eneric **SAS Tool**

**Shared G**eneric **SAS Tool**

21/05/2012
Agenda

- Introduction
- **Generic SAS Tool (GSAST) architecture**
  - Microdata processing
  - Architecture
- Metadata in GSAST
  - Structure
  - Management
- Possible extension of GSAST to data providers
  - Current processing
  - Self service
- Conclusions
Microdata Processing

Stage 1: Regular production chain

- Receipt of data
- Country processing
- Aggregation
- Disseminate

Member States

EDAMIS

Eurostat
Microdata Processing

Stage 1: Regular production chain

- Receipt of data
- Country processing
- Aggregation
- Disseminate

Data revision, justification

Revised data

Data quality checks

Error and quality report
Microdata Processing

Stage 1: Regular production chain

- Receipt of data
- Country processing
- Aggregation
- Disseminate

Calculation of European statistics; Disclosure control, etc.
Microdata Processing

Stage 1: Regular production chain

- Receipt of data
- Country processing
- Aggregation
- Dissemination

www.ec.europa.eu/eurostat

Anonymised microdata for researchers
Microdata Processing

Stage 1: Regular production chain

- Receipt of data
- Country processing
- Aggregation
- Disseminate

Stage 2: Additional ad-hoc processing
- further analyse the data
- provide statistics to answer external requests
- additional publications
GSAST, the corporate approach

- Requirements:
  - Use state-of-the-art technologies;
  - Apply a centralised approach;
  - Modular design including generic and reusable modules for all data collections;
  - Standard user interface;
  - Possibility to perform additional computations on the data (Stage 2);
  - Easy to use and easy to learn by statisticians;
  - Maintenance of the tool should be easy and could preferably be performed by the statisticians.
GSAST, the corporate approach

Implementation:
- SAS Business Intelligence platform: Server – Client approach;
- Common server-based tool with central development and support services;
- SAS Enterprise Guide (EG) as user interface with stored processes;
- Full computational power of SAS is available for additional processing;
- Easy to learn common user interfaces;
- Use of metadata to parameterise the production processes.
GSAST in Use

- GSAST is in use for 9 different surveys.
- In SAS EG the user initiates the operation one by one
  - Parameters are provided by users.
- Feedback is provided about the operations.

Status

<table>
<thead>
<tr>
<th>MESSAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update 2 records in CVT84GRA.TC8T84_DYNA_STADIUM_FILES</td>
</tr>
<tr>
<td>Workflow stadium_contents starts the 13JUL11 at 18:49:01 and finishes at 18:49:01 in 1 second</td>
</tr>
</tbody>
</table>
Agenda

- Introduction
- Generic SAS Tool (GSAST) architecture
  - Microdata processing
  - Architecture
- Metadata in GSAST
  - Structure
  - Management
- Possible extension of GSAST to data providers
  - Current processing
  - Self service
- Conclusions
**Metadata in GSAST: Structure**

- **Technical metadata**
  - Defines the computational environment
  - SAS BI architecture

- **Process definition metadata**
  - Input parameters
  - Process parameters

- **Structural metadata**
  - Datasets
  - Variables

- **Statistical metadata**
  - Flags
  - Footnotes
  - Code lists
Metadata in GSAST: Management

Maintenance of the surveys mainly by means of metadata updates

Complex metadata structure
- Integrity constraints
- Difficult to maintain manually
Metadata in GSAST: Management

GSAST Metadata Editor the maintenance tool

– Integrated to the existing GSAST as an EG add-in;
– User-friendly way to browse complex metadata structures;
– Parametric and metadata driven;
– Metadata editing and validation to maintain integrity;
– Centralised management of code lists by the SDMX registry.
Agenda

- Introduction
- Generic SAS Tool (GSAST) architecture
  - Microdata processing
  - Architecture
- Metadata in GSAST
  - Structure
  - Management
- Possible extension of GSAST to data providers
  - Current processing
  - Self service
- Conclusions
GSAST extension for data providers

- Well-established workflows for microdata collections at NSIs.
- Data Quality checks at Eurostat at different levels.
- There is an overlap between the processing workflows at the NSIs and at Eurostat.
- Provide remote access to NSIs for the Country processing in the GSAST chain.
- Self-Service approach.
GSAST: Current processing

1. Data Upload
2. Country Processing
3. Feedback on the data
4. Data Upload
5. Country Processing
6. Multiple Countries Processing

EDAMIS

GSAST

Country Responsibility

Eurostat Responsibility

Country Loop
GSAST: Proposed workflow

1. Data Upload
2. Country Processing
3. Feedback on the data
4. Data Upload
5. Country Processing
6. Multiple Countries Processing

Country Responsibility
Eurostat Responsibility

GSAST 1: Web Based Interface

GSAST 2
Agenda

- Introduction
- Generic SAS Tool (GSAST) architecture
  - Microdata processing
  - Architecture
- Metadata in GSAST
  - Structure
  - Management
- Possible extension of GSAST to data providers
  - Current processing
  - Self service
- Conclusions
Conclusions:

– The GSAST platform is proven to be useful to process of several microdata collections.
– GSAST Metadata Editor solved the problem of the relatively heavy metadata for survey maintenance.
– Possible extension towards NSIs presents several challenges.
– While the proposed architecture ensures secure treatment of microdata, the adaptation of working methods presents a major challenge which is, however, fully in line with the joint strategy of the ESS.