# » Statistics Portugal Survey Management System Architecture and SIMSTAT

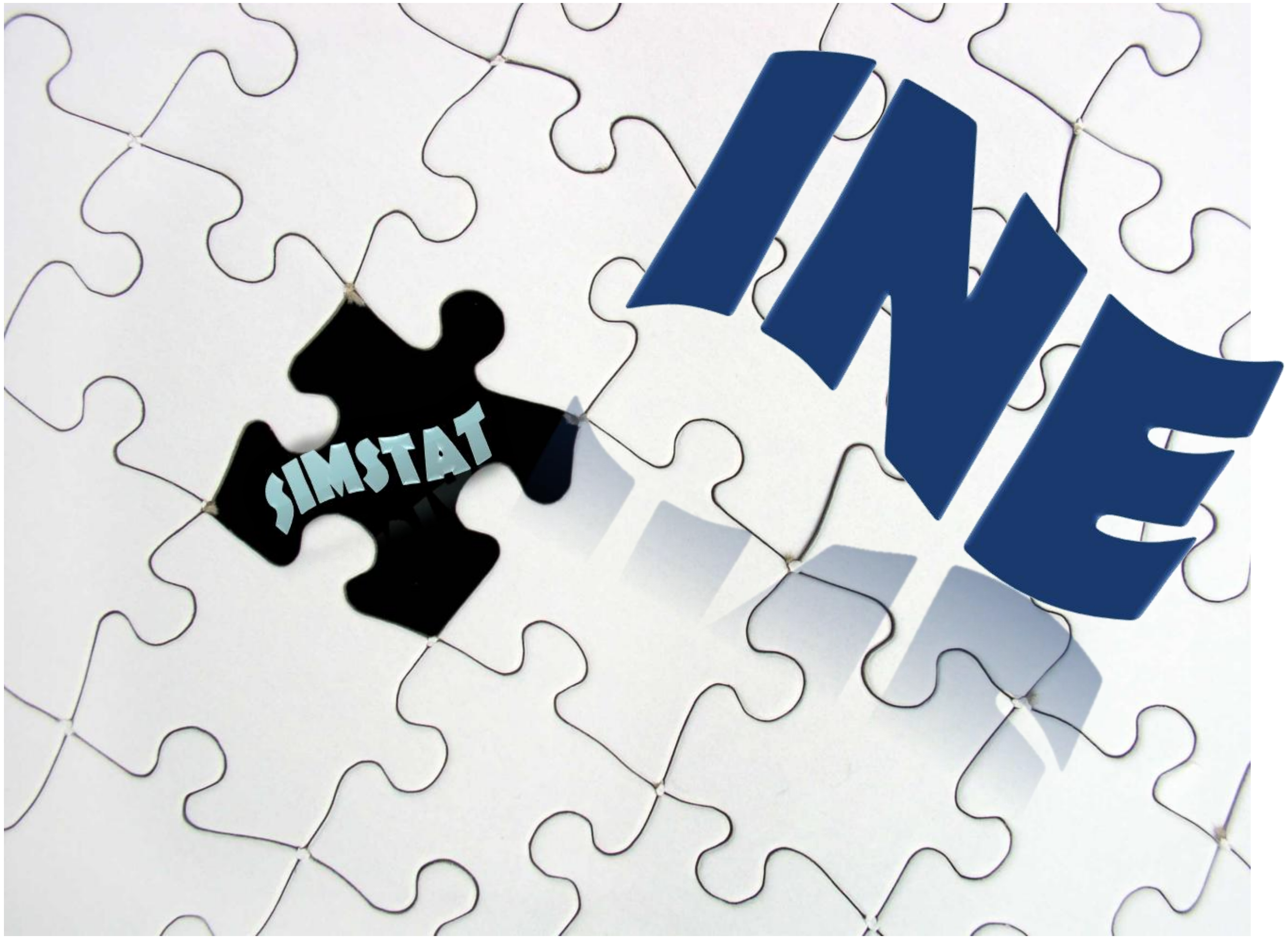
TF - 7th ESS. VIP IT «
**Luxembourg**

Duarte Bruno

**5-6th November 2014**

»

# SIMSTAT@INE



# Survey Management System
# Business Statistics

- Objectives

  – Implement an integrated and consistent approach to survey management systems

  – Integrate, harmonize and reuse processes

  – Increase coherence and comparability by using the same updated population

  – Reduce statistical burden on data providers

# Survey Management System Subsystems


# Integrated Information System

# Survey Management System

Fue Sigua Gresp GPap Webreg webinq SICC

- Business Register
  - Statistical Unit Registers System for business
    - Business; Local Unit; Road Vehicle
  - Refreshed via administrative sources and surveys
  - Online shared data (access and update)
  - Definition of updating rules for statistical units

# Survey Management System

Fue Sigua Gresp GPap Webreg webinq SICC

- Population and Sample Management
  - Population, sampling frame and sample central database
  - Focus on sample design and management
  - Trace down the statistical units life cycle
  - Create conditions to facilitate burden measurement

# Survey Management System

Fue Sigua Gresp GPap Webreg webinq SICC

- Respondents Management
  - Optimize cooperation and encourage participation of respondents
  - Respondents database (Webinq members and non members)
  - Online shared data (access and update)
  - Definition of updating rules for respondents data

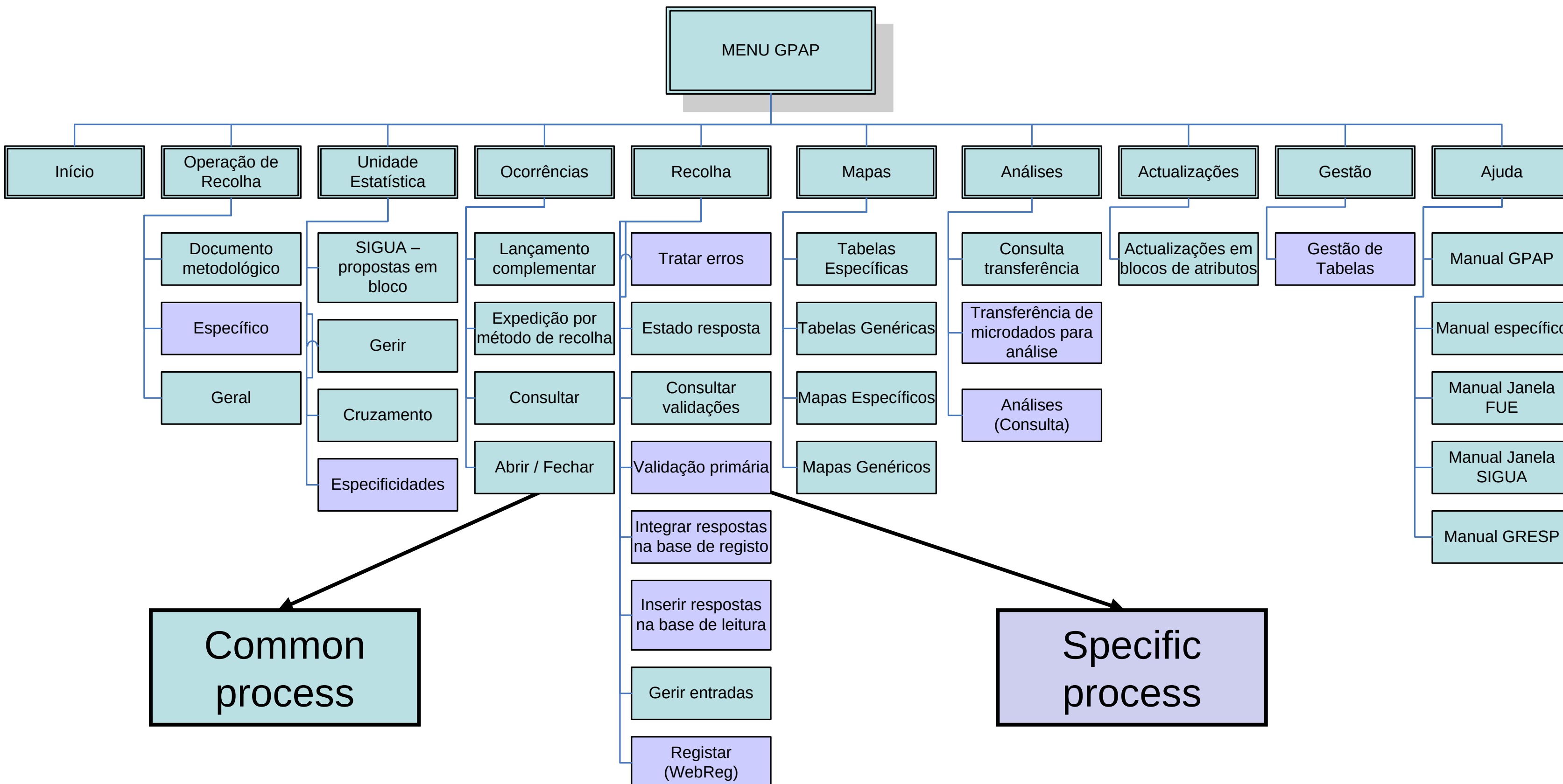
# Survey Management System

Fue Sigua Gresp GPap Webreg webinq SICC

- Data Collection Process Management
  - Intercommunication with:
    - Business Register (FUE)
    - Samples (SIGUA)
    - Respondents (GRESP)
  - Execution and management of:
    - Common and standardized processes
    - Specific processes - according to each survey characteristics
  - Transmission of validated micro-data to the Data Warehouse

# Survey parameterization

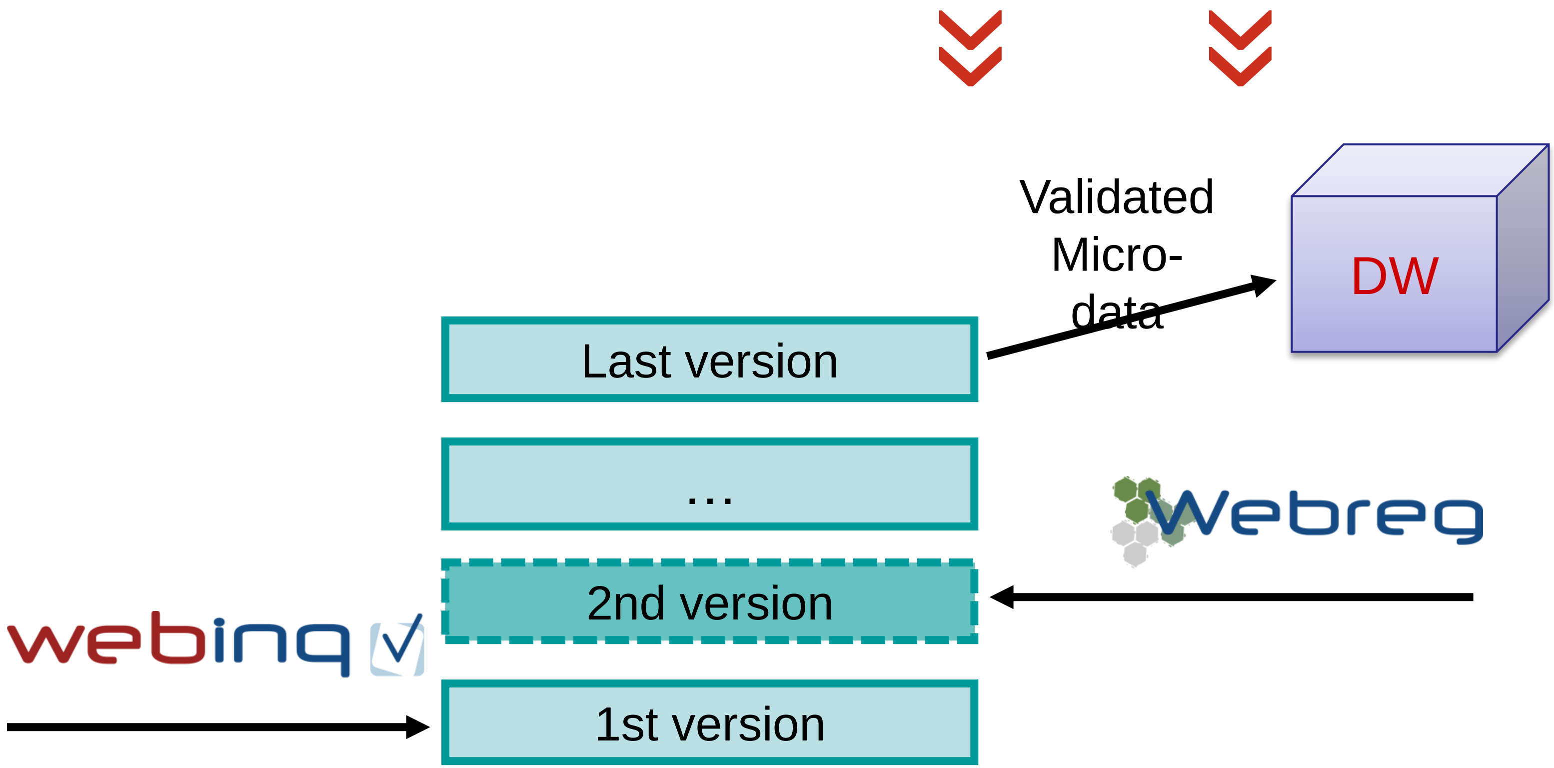
Fue Sigua Gresp GPap Webreg webinq SICC



# Survey Management System

- Data Capture and Editing
    - Internet (Online responses)
    - Inside INE (Paper responses post-collection and editing)
  - Data Version Stack – Four main rules:
    - **Single** response for each tuple (survey, data reference period, statistical unit)
    - Data editing performs a **new** response version
    - **All** versions of the responses are saved
    - **Last** version goes to the data warehouse.

# Data Version Stack



# Survey Management System

Fue Sigua Gresp GPap Webreg webinq SICC

- Call Centre
  - Last system to be implemented
  - Integrated call center
    - Telephone
    - Mail
  - Management of respondents interactions
    - Inbound
    - Outbound
  - Interactions are context aware and tracked

# Survey Management System

100%

**2008**

12 surveys
51 464 respondents
89 081 business
761 818 responses

**2010**

26 surveys
85 583 respondents
125 679 business
1 768 444 responses

**2012**

55 surveys
113 436 respondents
161 810 business
3 240 163 responses

**2014**

74 surveys
144 825 respondents
193 780 business
> 4 700 000 responses

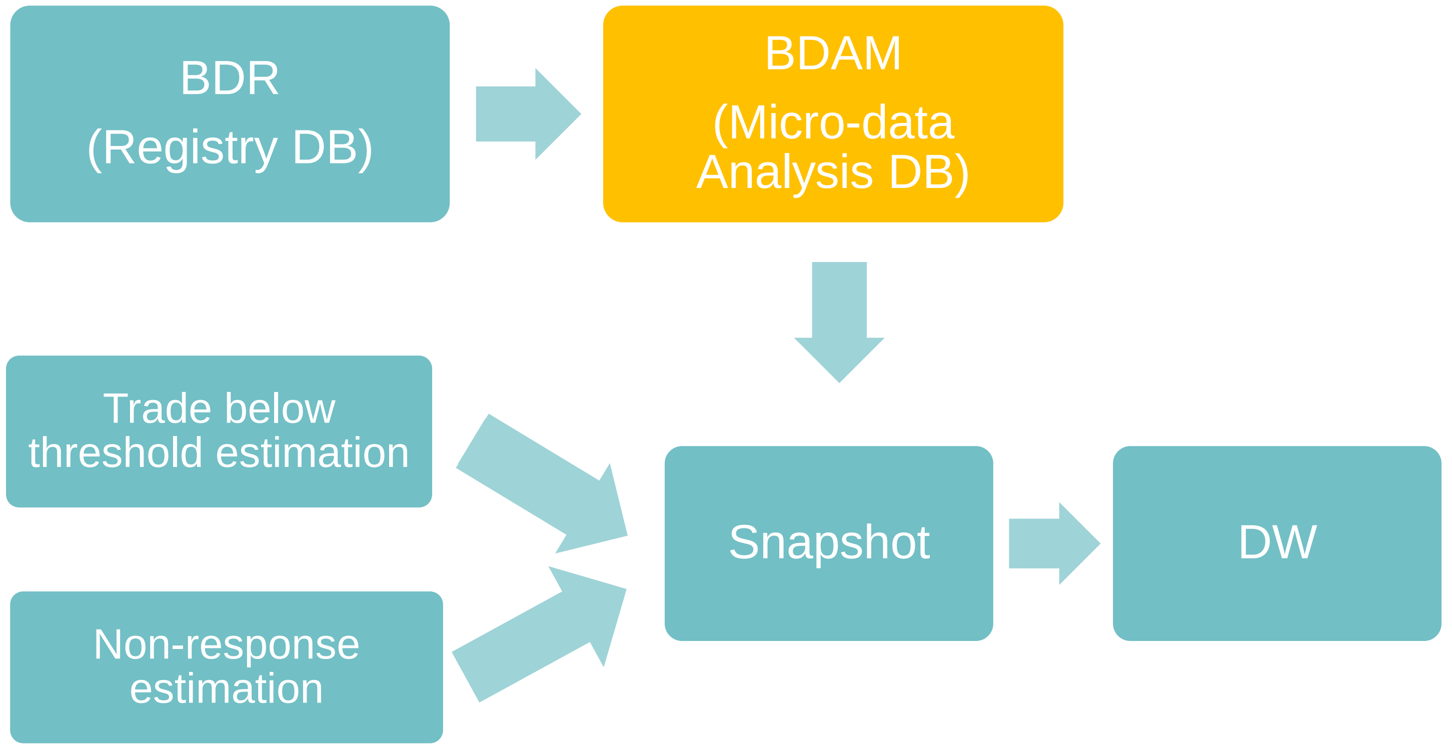
**Accumulated values**

# Integrate a survey (Requirements)

- Create in Metadata System (SMI)
- Register data sampling method (SIGUA)
- Parameterization (SIGUA and GPAP)
- Identify common processes to be used
- Prepare specifications and code specific processes:
  - Data capture and editing (Electronic form and Validation rules)
  - Imputation
  - Specific reports
- Guide to integrate surveys

SIMSTAT


# Intrastat Data Flow



# SIMSTAT - Micro-data

- Decisions and challenges

Tuple identification in BDAM
Defining (qualifying) vs non-defining (quantifying) values in PSI level response

Tuple life-cycle management
Identify modified micro-data from PSI responses
Generate a balanced differential list

Upload data to the CCN/CSI Hub
Synchronizing SIMSTAT to latest BDAM Snapshot

# Tuple definition

## Qualifying values

**Less likely to change**

- Sending MS
- Receiving MS
- Reference period
- Flow
- PSI_Id
- Partner ID
- Commodity code
- Country of origin
- Nature of transaction
- Delivery terms
- Mode of transport

**URI defining**

---

**More likely to change**

- Quantity in Net Mass
- Quantity in Supplementary Unit
- Taxable Amount
- Statistical Amount

**URI non defining**

## Quantifying values

Time=1

## Initial snapshot

Time=n

Identify differences and difference types to previous snapshot

# Difference map



- SIMSTAT's response data vs BDAM's

  1 - Same tuple (nothing changed)
  → *Ignore*

  2 - New tuple
  → ***I**nsert (new URI)*

  3 - Tuple was deleted
  → ***D**elete (old URI)*

  4 - Same tuple (quantifying value changed)
  → ***D*** + ***I*** *(old URI)*

  5 - Changed tuple (defining value changed)
  → ***D*** *(old URI)* + ***I*** *(new URI)*

# Change detector at micro-data level (tuple)

## Response #1

**a)** New micro-data

Tuple ∈ BDAM &

Tuple ∉ SIMSTAT

## Response #n

**c)** Inserted micro-data

Tuple ∈ BDAM

Tuple ∉ SIMSTAT

**b)** Deleted micro-data

Tuple ∈ SIMSTAT &

Tuple ∉ BDAM

**d)** Updated micro-data

Tuple ∈ (BDAM ∩ SIMSTAT) &

Values(BDAM) ≠ Values(SIMSTAT)

## Action

*Insert (new URI)*

*Delete (old URI)*

*Insert (old URI)*

# Technical Approach

**Response (BDR)**

- responseID
- firstResponseID
- Response data
- + sendjobID

**Micro-data (BDAM)**

- Micro-data
- + URI
- + sendjobID

INTRASTAT (existing tables)

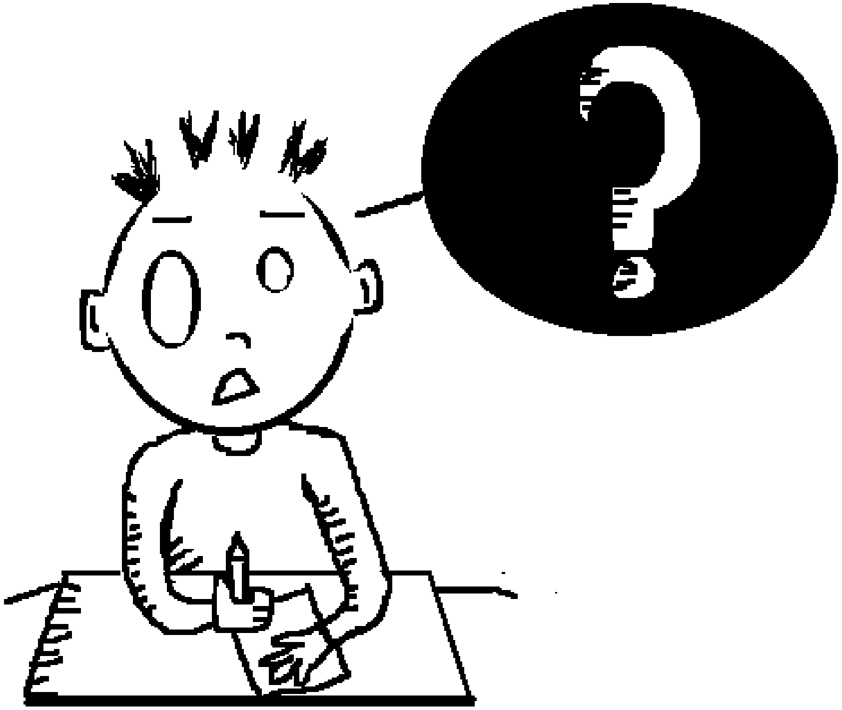
SIMSTAT (new table)


**+ Snapshot (SIMSTAT)**

- responseID
- firstResponseID
- Micro-data
- URI
- sendjobID
- status
- isCurrent
- auditing fields

Upload to HUB

# Questions



QUESTIONS

# » Statistics Portugal Survey Management System Architecture

TF - 7th ESS. VIP IT «
**Luxembourg**

Merci!
Thank you!
Obrigado!

Duarte Bruno– duarte.bruno@ine.pt

**5-6th November 2014**

»