



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
Workshop on Statistical Data Collection 'Resourceful Data Acquisition'
10-12 October 2018, Geneva, Switzerland



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Business Data Collection: From questionnaires to data sets

Paulo Saraiva
paulo.saraiva@ine.pt

What are we going to talk about?

- Business Data Collection: Where we are?
- Discuss automated business data interchange.
- Sharing the experience of Portugal in this field.
- Pointing-out some challenges.

References: JOS Paper (2018)

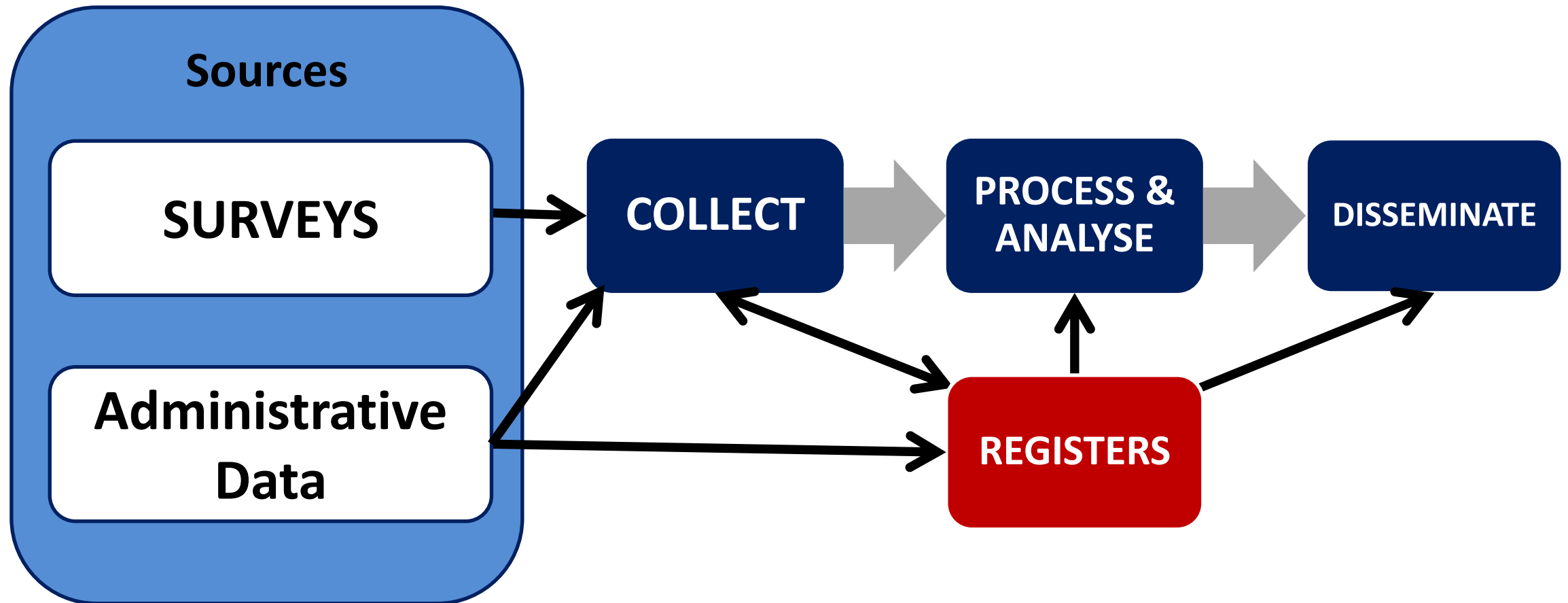


Journal of Official Statistics, Vol. 34, No. 2, 2018, pp. 419–443, <http://dx.doi.org/10.2478/JOS-2018-0019>

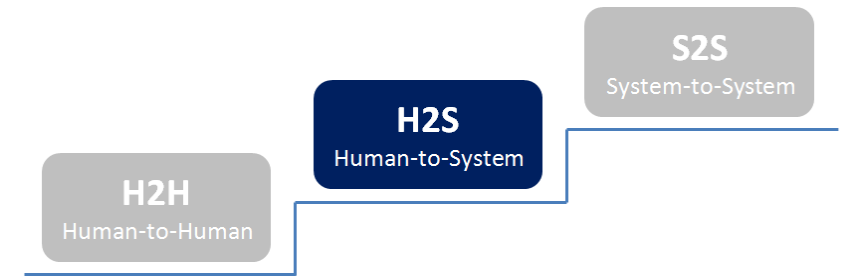
Business Data Collection: Toward Electronic Data Interchange. Experiences in Portugal, Canada, Sweden, and the Netherlands with EDI

*Gert Buiten¹, Ger Snijkers², Paulo Saraiva³, Johan Erikson⁴, Anna-Greta Erikson⁴,
and Alice Born⁵*

Business Statistics: Production Cycle



Let's go 'all-electronic'



- Statistical Producers move to web-based data collection.
- In 2005, Statistics Portugal introduced integrated web questionnaires as an option for business surveys.
- In Portugal, like the other countries, all business surveys use web questionnaires nowadays.
- 98% of the collected questionnaires (2017)

webinq

Administrative Sources and Big data

- Registers and **administrative data** were being used.
- Replacing surveys, supplementing data, validation or imputation.
- Another secondary data source gets attention is **big data** (since 2013).
- We are working towards both the extended use of secondary sources (including big data) as well as of web surveys.
- **Multi-source approaches combining primary and secondary data.**
- This process is still going on.

The way to Integration

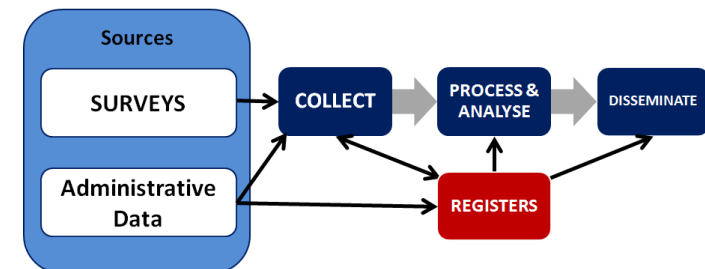


■ Conceptual **harmonisation**:

- Eliminating differences in definitions, concepts, classifications, variables...
- Sharing same metadata and taxonomies.
- Both inside (horizontal) and outside (vertical) the organization.

■ Technical **standardisation**:

- Uniform data formats, file structures, transmission modes, software



Harmonisation: Metadata

- Metadata provide information on data, and describe them: they are the data definitions (UNECE 1995).
- Essential in an integrated statistical process.
- Metadata support the whole statistical production.
- Concepts, classifications, variables, data sources, and outputs.
- They are as important as data.
- A core tool to a Statistical Producer.

Portugal: Integrated Metadata System

<http://smi.ine.pt>

- Built in 2003, it covers all Portuguese Statistical System
- Repository of harmonized concepts, classifications, variables, data collection instruments and methodological documentation.
- Its purpose is to support:
 - Survey design;
 - Data outputs.

Interface: Contents Tem conta? [Login](#)

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Metadata System

Enter a keyword [Advanced Search](#)

General Concepts Classifications Methodological documentation Variables Data collection instruments

SMI

Concepts

Module that integrates and provides terms and definitions of concepts used for statistical purposes on the National Statistical System, which underlie the consistency and comparability of the data collected and disseminated. [Read more ...](#)

Code	Designation
2051	ECONOMIC ACTIVITY

Metadata System

The Metadata System integrates and provides concepts, classifications, variables, data collection instruments and methodological documentation in the scope of the the National Statistical System (NSS). The various components of the system are interrelated, aim to support statistical production and document the dissemination of Official Statistics. The Metadata System is a coordination and harmonization tool within the NSS. [Read more ...](#)

Methodological Documentation

This module includes and provides access to documents (called Methodological Documents) that describe, at present time, each statistical operation undertaken under the National Statistical System. It respects a common structure with predefined topics. [Read more ...](#)

Code	Designation
270	Survey on ICT usage in enterprises

Variables

Module that integrates and provides the observed variables and statistical indicators disseminated by the National Statistical System. [Read more ...](#)

Code	Designation
10009	Compensations (€) of Enterprises by Ec...
11900	Live births (No.) by Place of residence o...

Classifications

Module that integrates and provides national, Community and international classifications used for statistical purposes on the National Statistical System. [Read more ...](#)

Code	Designation	Acronym
V00554	Portuguese classification of economic activities, revision 3	CAE Rev. 3
V00017	Code of the administrative division	
V03503	Nomenclature of territorial units for statistics, 2013 version	NUTS 2013

Data Collection Instruments

This module includes and provides information concerning data collection instruments registered at INE, and access to their image. Whether they may be questionnaires (when connected to surveys) or files (in electronic format). [Read more ...](#)

Code	Designation
10053	IUTICE - INQUÉRITO À UTILIZAÇÃO D...

Portugal: Integrated Survey Management System

- Statistics Portugal, in 2005 a central data collection department was established.
- Integrated Survey Management System (called SIGINQ) in 2007.
- First SIGINQ covered the business surveys, and followed by the agriculture and household surveys.



Portugal: Integrated Survey Management System

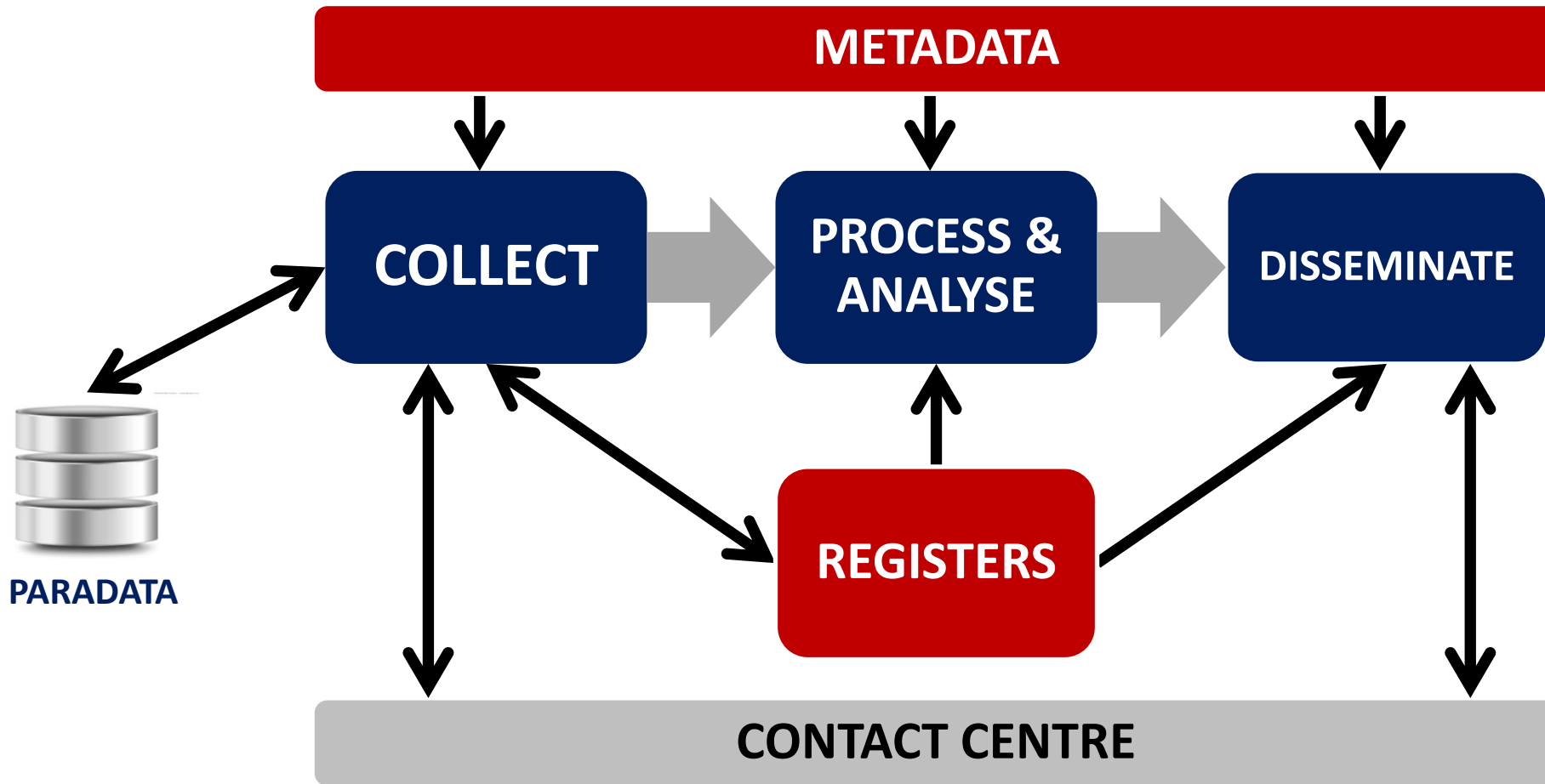
- Integration in this system is based on four pillars:
 - processes, metadata, registers, and variable-oriented.

SIGINQ



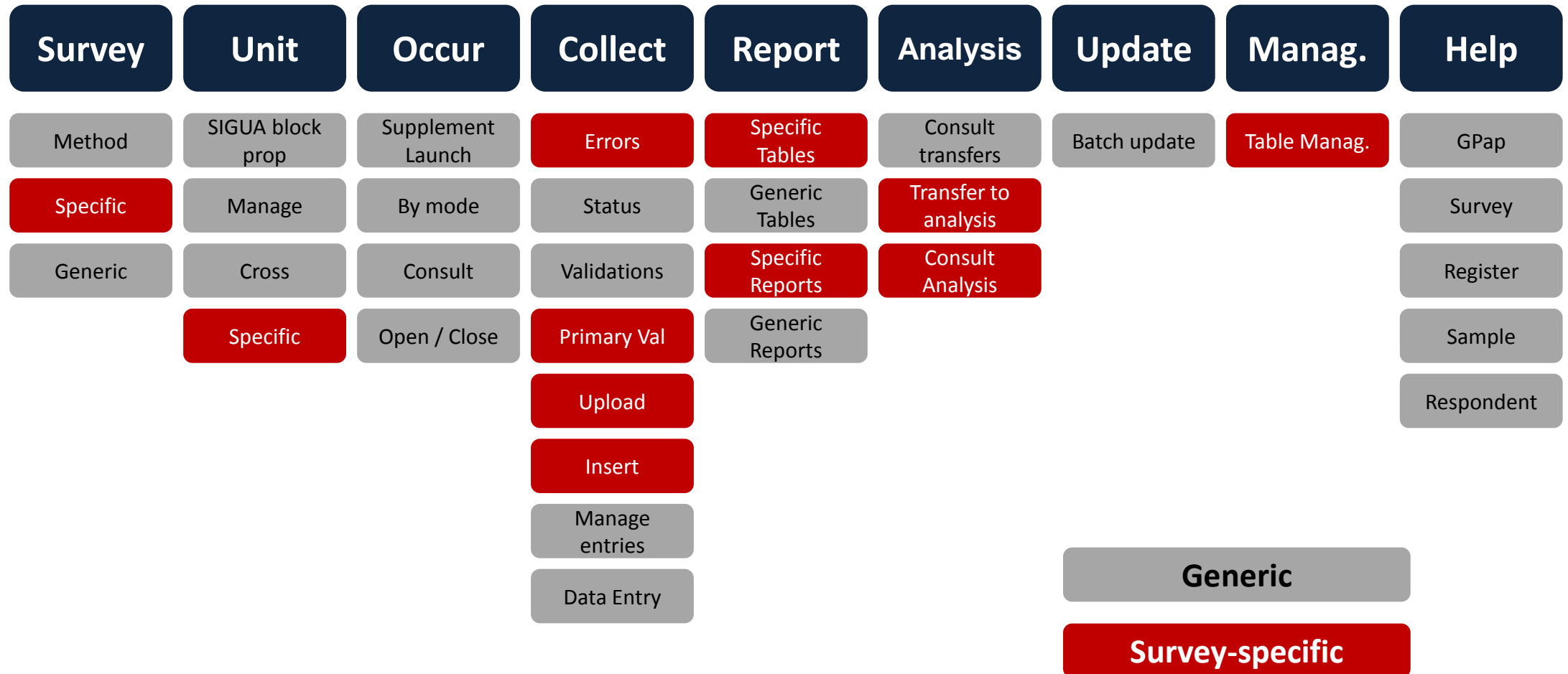
Portugal: Integrated Survey Management System

- Following the principle of maximizing the reuse of components, the system combines generic functions with survey-specific features.
- Some components are shared by all surveys (business or household), and others are specific for a single survey.
- Sometimes, a specific component is adapted to become a generic one, if it has a certain potential.



Reuse of components

SIGINO



Process Management

SIGINQ

GPap v4.0.2016.0 de 11-05-2017







SIGINQIAP | Surveys | Reports | Analysis | Management | Help

PT 987 654 321 - Company Name

Filtro

Survey

Show 25 Procurar:

	Survey	Reference	Activation	Due Date	Mode	Date (Respondent ID)	Type
	IACC	201700	01-02-2018	19-03-2018	WebInq	09-03-2018 (CFR000110)	Resposta
	IACC	201400	05-06-2015	22-06-2015	WebInq	17-06-2015 (PCL000158)	Resposta
 	ICVG	201700	20-04-2018	08-05-2018	WebInq	07-05-2018 (CFR000110)	Resposta
 	ITENF	201704	05-01-2018	25-01-2018	WebInq	09-02-2018 (CMC000202)	Resposta

PDF View

Edit Function

Standardisation in Portugal

- In many countries, technical standards are not regulated by law.
- Portugal chose for another solution.
- In 2007: Simplified Business Information system (IES) – legal act.
- A mandatory nation-wide initiative to enable S2S financial reporting between businesses and government agencies.



Diário da República, 1.ª série — N.º 12 — 17 de Janeiro de 2007

MINISTÉRIO DA JUSTIÇA

Decreto-Lei n.º 8/2007

de 17 de Janeiro

Statistics Portugal's Automated Data Transmission



- In addition to web questionnaires, businesses may also use another data collection mode: Automated Data Transmission (TAD).
- It is a S2S approach with the upload of **XML files**, or a dedicated Web service.
- TAD allows businesses respondents to prepare an XML envelope, with variables from **one or more companies and surveys**, avoiding the completion of e-questionnaires.
- After uploading the file, it integrates the variables accordingly, which can be accessed by the respondent or by data collection staff.
- Currently available for 14 surveys.

TAD


Transferência Automática de Dados


BUSINESS




- After requesting to use this service, a Globally Unique Identifier (GUID) code is generated to the respondent, which provides access to the TAD service.
- Technical guides are available.
- Formats and instructions on how to integrate in external applications,
- and the use of web services for an automatic and unattended solution.


Technical Guides

 INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

 XML files upload

GENERIC GUIDE

 INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

 XML files upload




SURVEY GUIDE

IPHH

 INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

 XML files upload

INTEGRATION GUIDE

-  VALIDATION RULES
-  VARIABLES
-  Examples



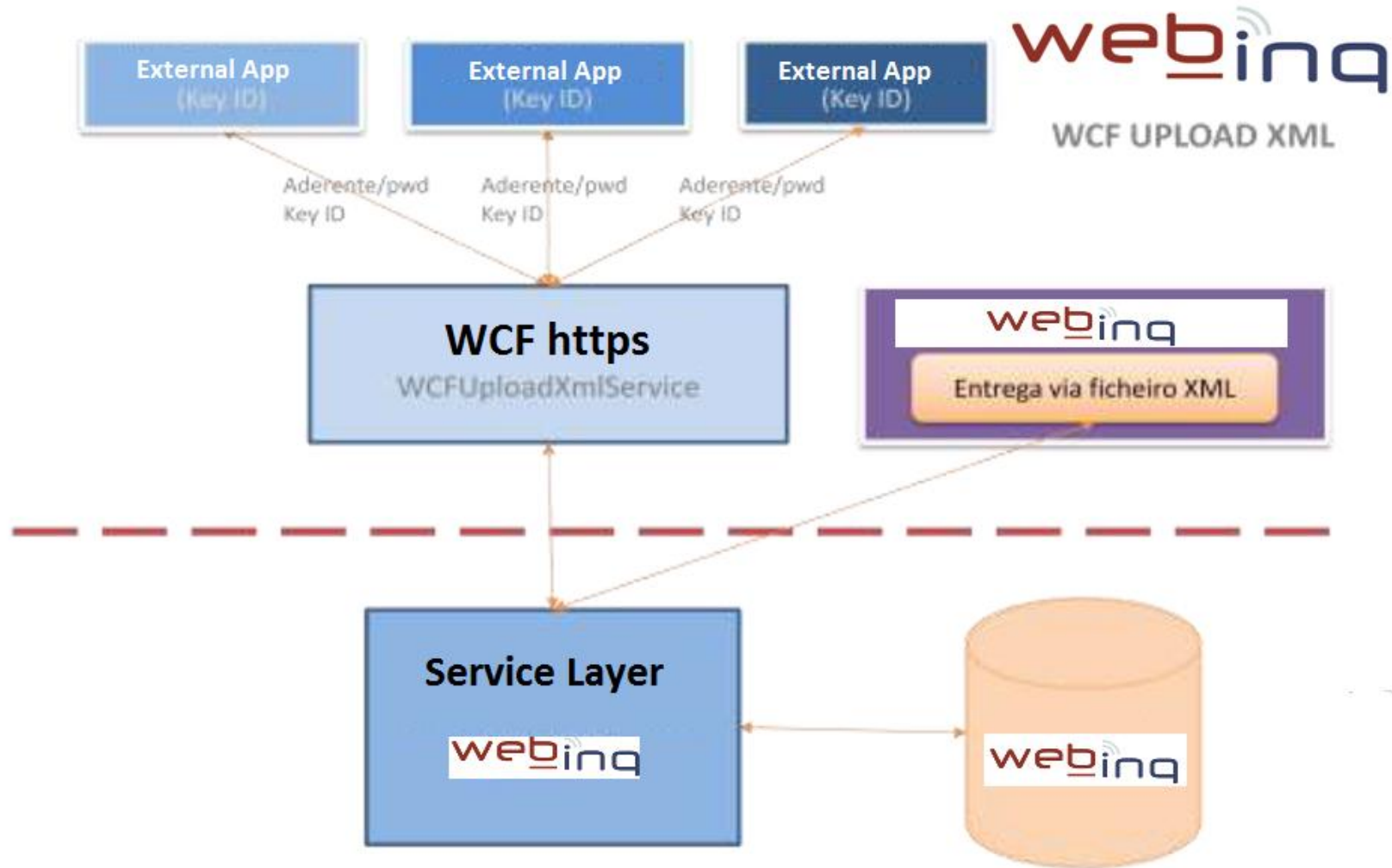
Basic XML Structure

```
<?xml version="1.0" encoding="utf-8" ?>
<ine xmlns="http://ine.pt/public/xsd/forms"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://ine.pt/public/xsd/forms upload_v1.xsd"
>
  <header chave="...">
    ...
  </header>
  <respostas>
    <resposta>
      <referencia>
        ...
      </referencia>
      <dados tipo="..." ult_atualizacao="YYYY-MM-DDTHH:MI:SS">
        ...
      </dados>
      <info>
        ...
      </info>
    </resposta>
  </ine>
```

Variables description for a Survey

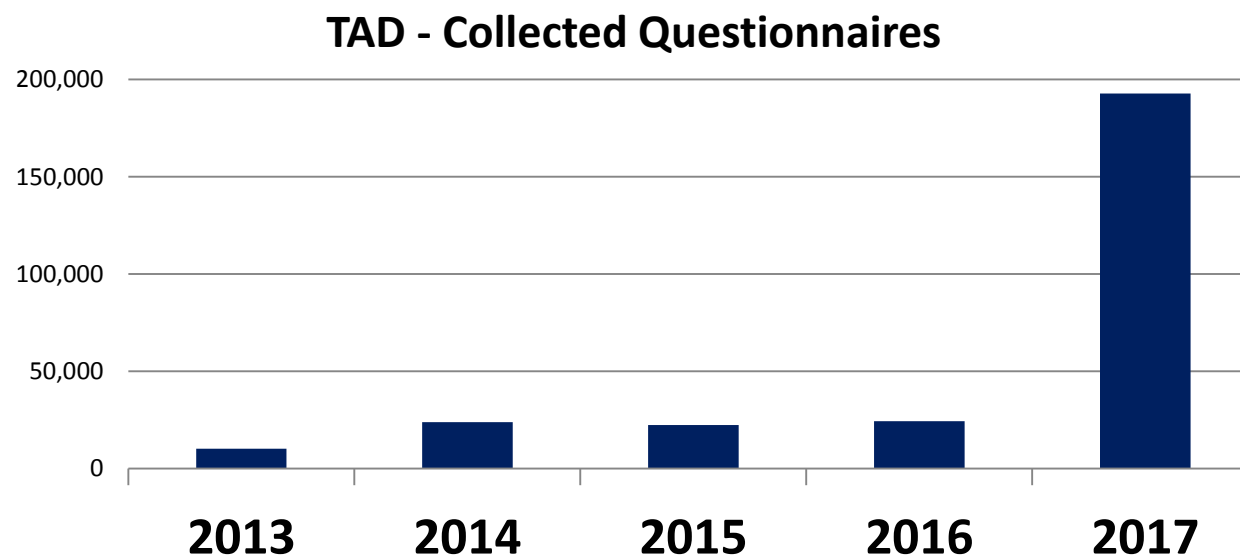
Variable ID	Description	Type	Fomat	Domain	Segment
BC101	Economic Activity	num	5	list:CAE	d
BC105	Activity Status	num	2	list:STA	d
BC151	Sazonality?	num	1	list:yes_no	d
BC153	Establishment Category	txt	2	list:categories	d
BC156	Beach	num	8	lista:beaches	d
V100	Number of available rooms	num	5	>=0	c
V101	Total of single beds	num	5	>=0	c
V102	Total of double beds	num	5	>=0	c
V103	Total of occupied rooms	num	6	>=0	a

Integration: Application Guide

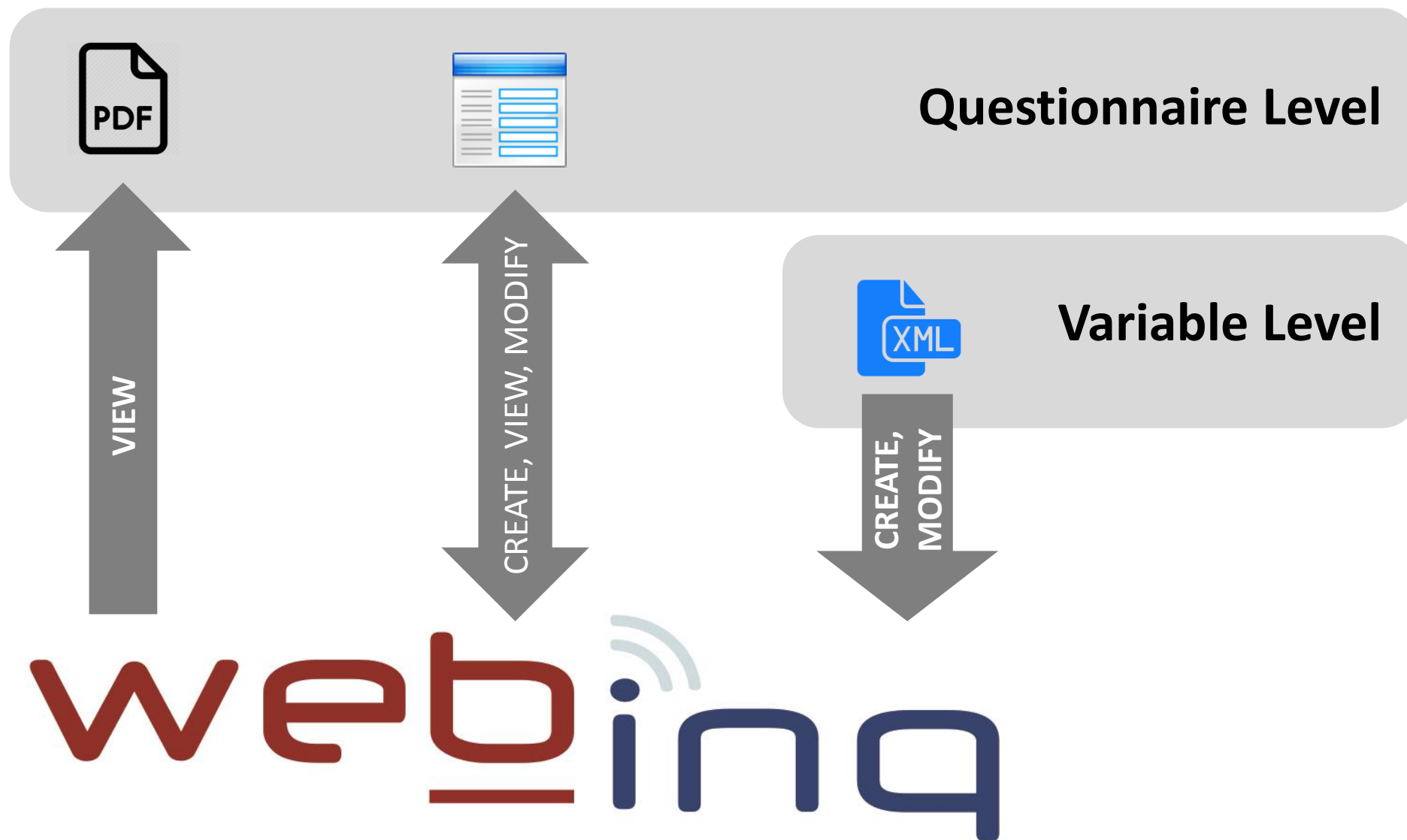


TAD acceptance and the way forward

- **24% of the collected questionnaires (WebInq, 2017)**
- There has been an increasing acceptance, indicating a significant reduction of the resources involved, as well as increased quality.



Multiple ways to create, view and modify



Feeding back data to businesses in Portugal

- Statistics Portugal evaluated possibilities of including information flows from the users back to the suppliers.
- Providing feedback and benchmark data is an incentive for businesses to participate in surveys, even if mandatory.
- It may serve as a driver for businesses, their service providers and software developers to adopt S2S.
- Such ‘feedback data’ can be used by them to provide better and even new services to their customers.

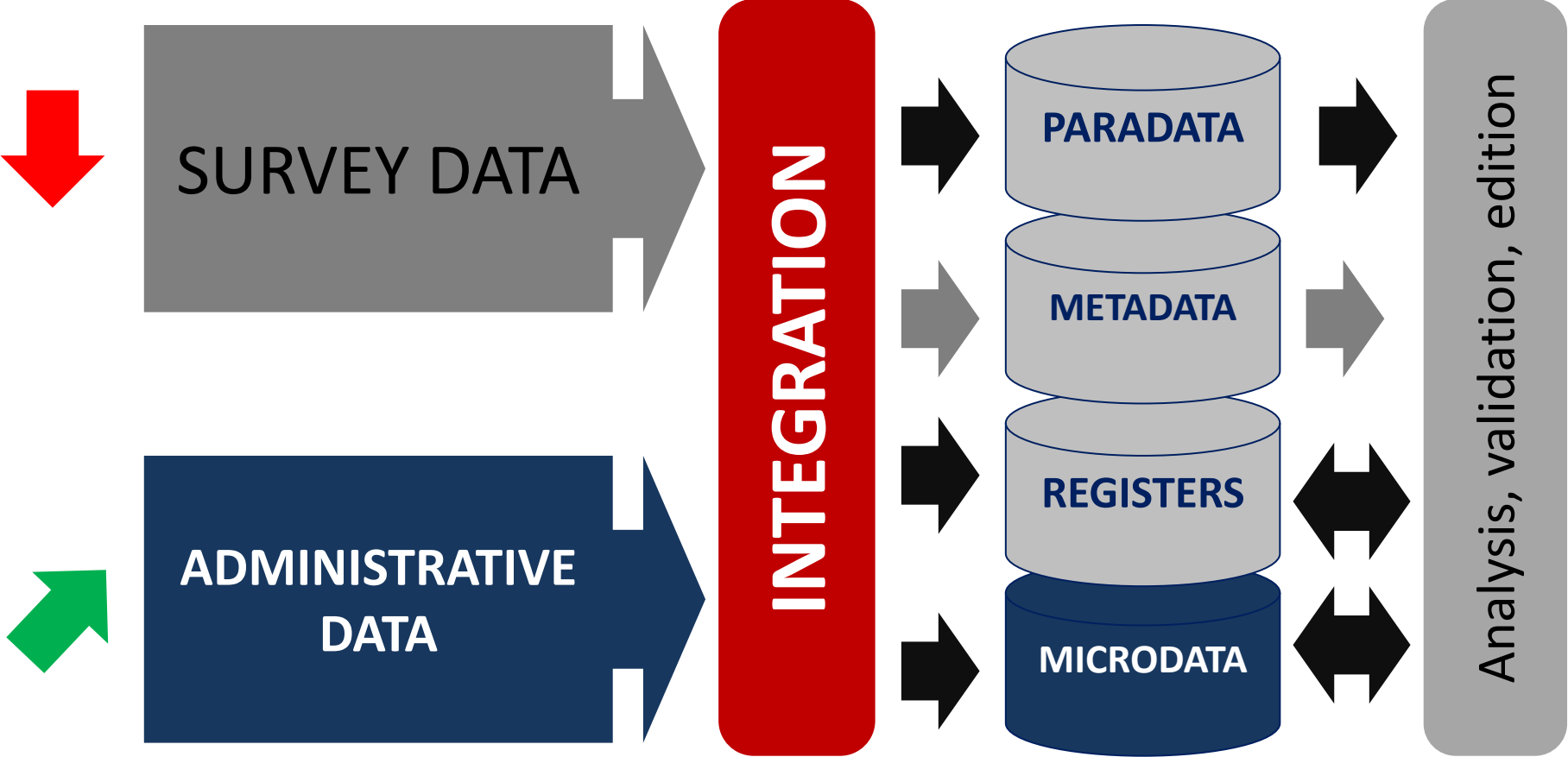
Challenges



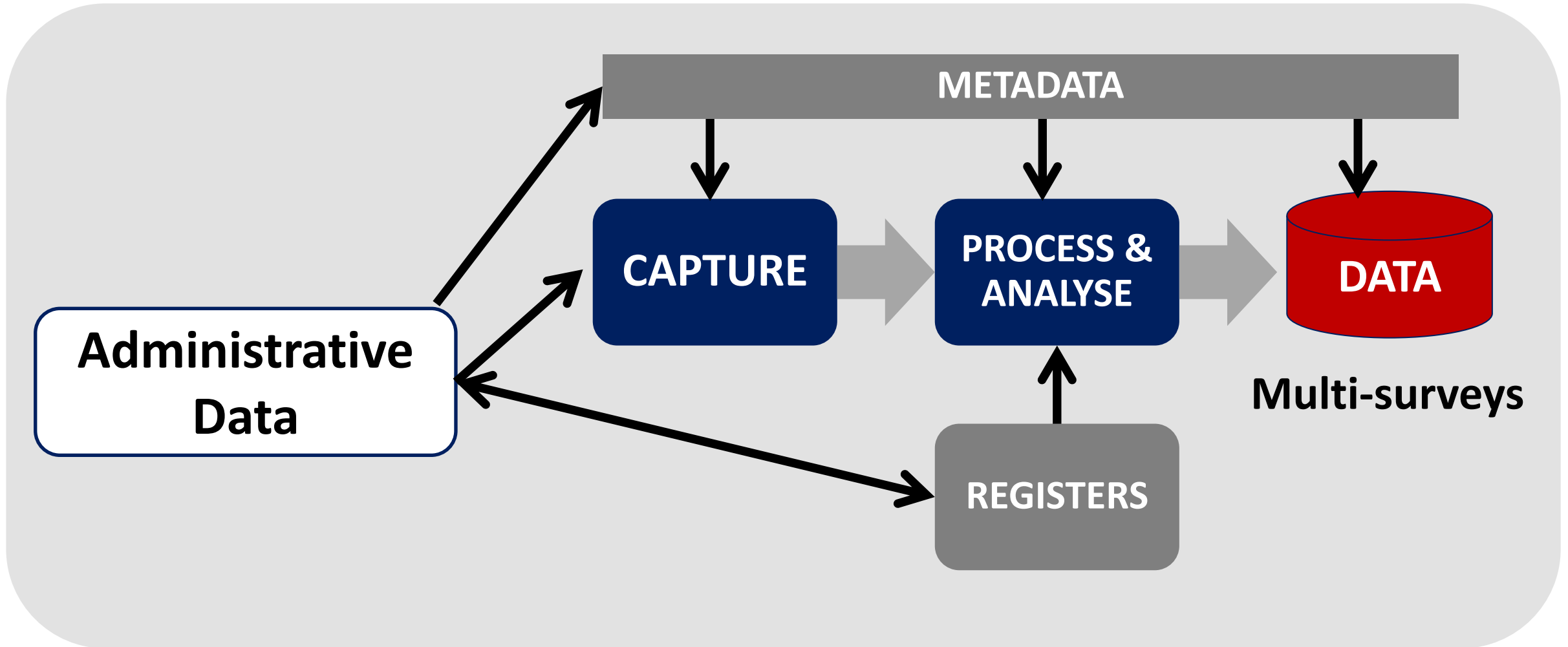
Challenges: Methodological Approaches

- Sampling has been an efficient way to collect data from a population of units.
- **Is sampling still efficient within an S2S approach?**
- Maybe a census is much more efficient, considering the required investments by businesses, software developers.
- This applies to the IES system, which includes all businesses in Portugal that by law need to report financial statements to the government.

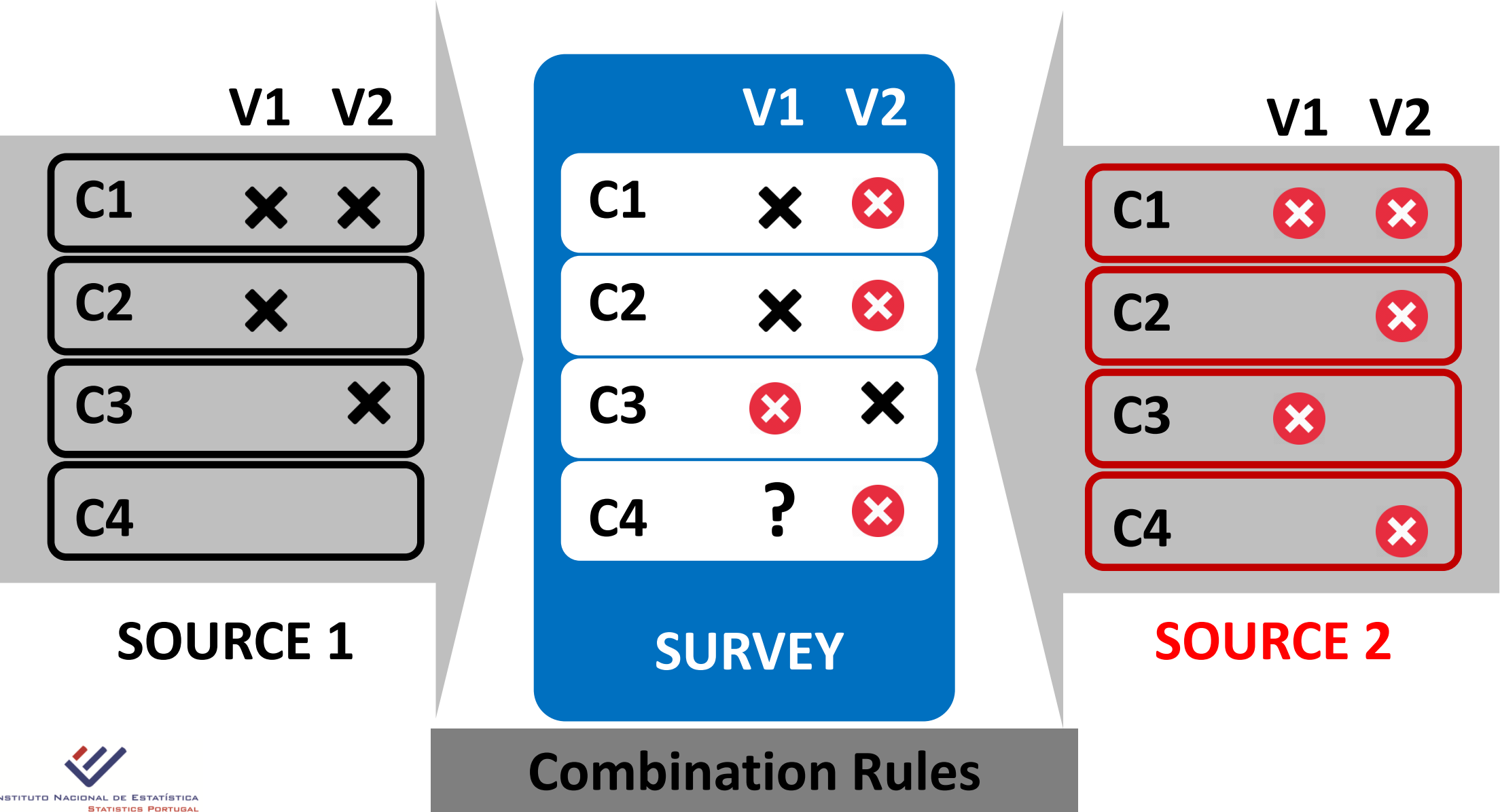
Challenge: Source Integration



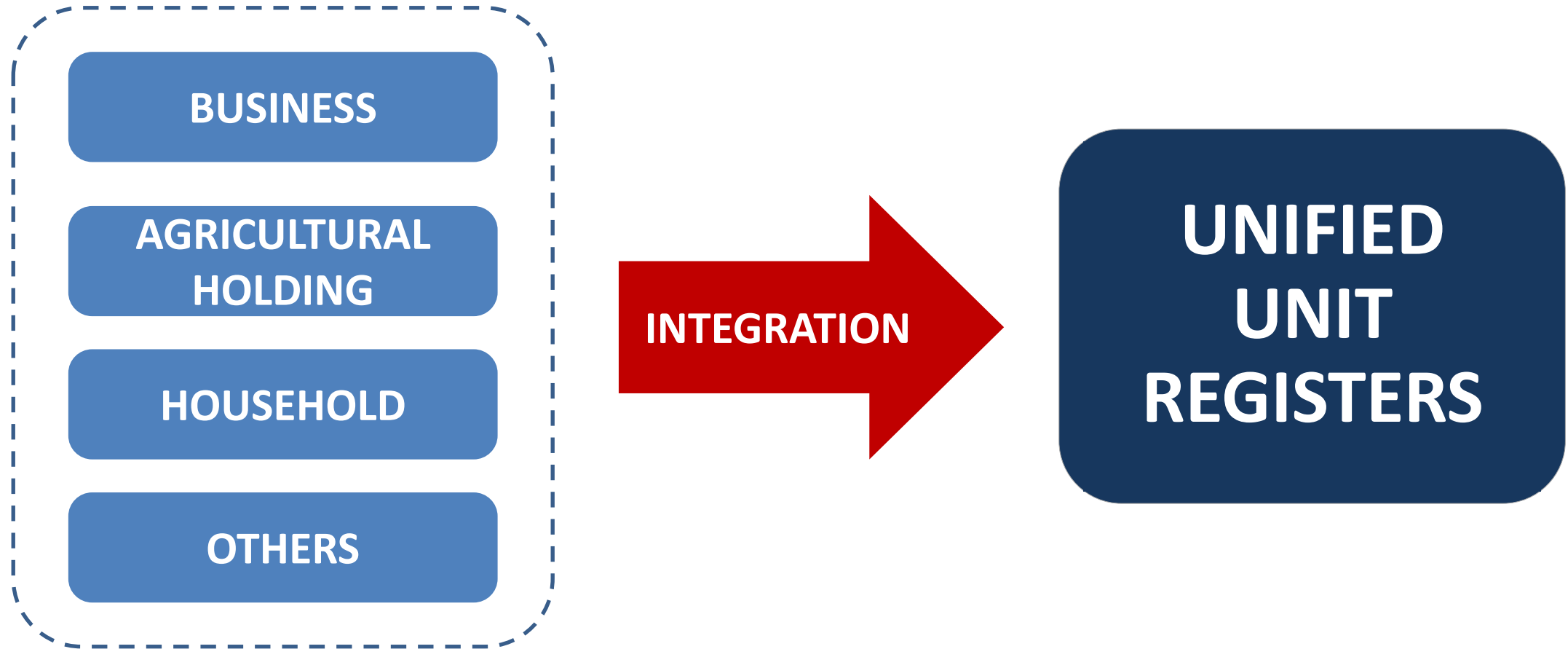
Challenge: Administrative Sources – Processes



Challenge: Data Integration

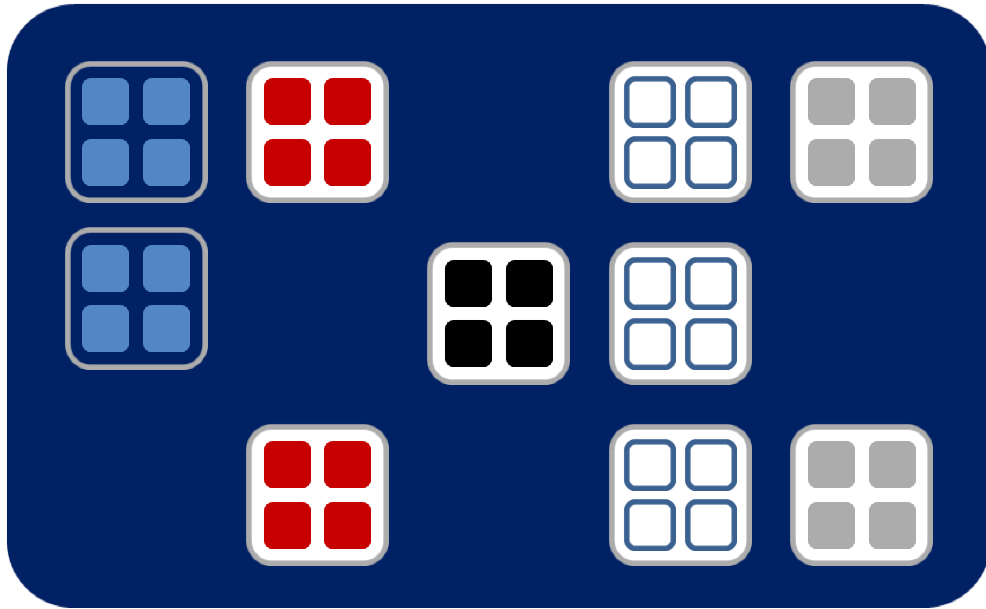


Challenge: Unified Unit Registers

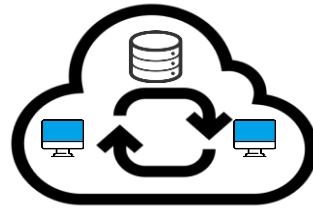


Challenge: Reengineering Surveys

Questionnaires



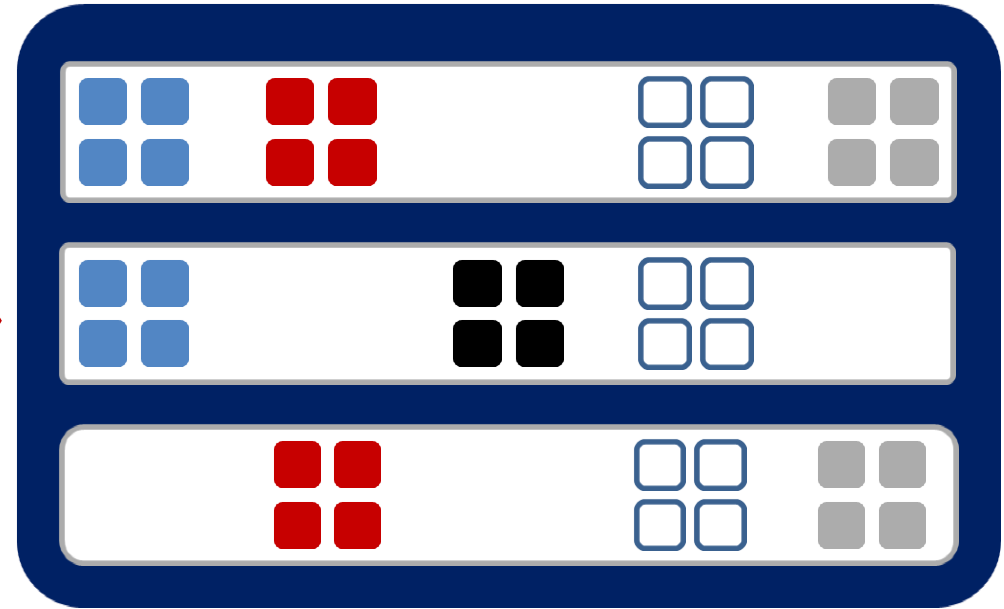
Proliferation of questionnaires
and surveys



S2S

**REENGINEER
SURVEYS**

Data Sets



One Data Set for several Surveys

Business Data Collection: From questionnaires to data sets

thank you!

Paulo Saraiva
paulo.saraiva@ine.pt