



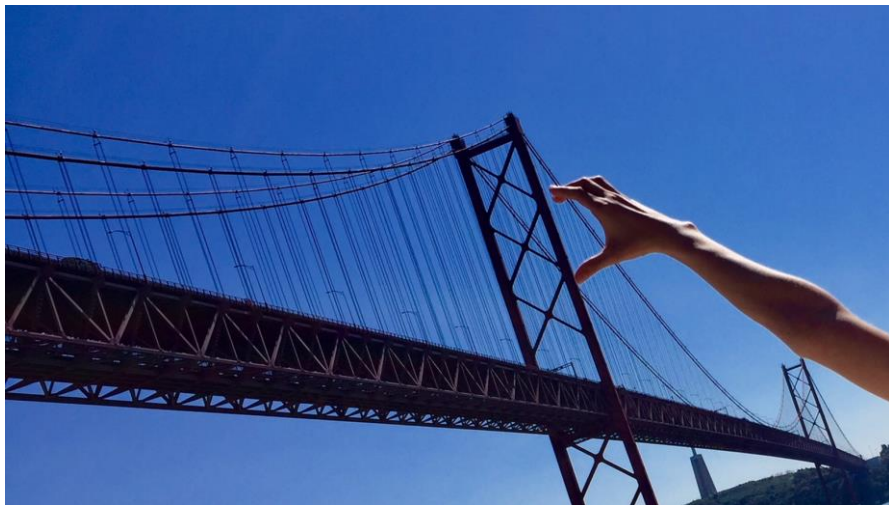
UNECE

modernstats
an HLG - MOS



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Workshop on Statistical Data Collection 2019
14-16 October 2019, Geneva



Multiple Sources in Business Surveys

Paulo Saraiva

paulo.saraiva@ine.pt

Almiro Moreira

almiro.moreira@ine.pt

Sofia Rodrigues

sofia.rodrigues@ine.pt

Statistics Portugal and Administrative Sources

- Already used in several areas of official Portuguese statistical production:
 - **+100 sources** benefiting **80 statistical activities**.
 - Business and Social statistics
 - Data Warehouse with **37 regular flows** and 1.5 billion historical observations.

Tax and Customs Authority as one major Data Source

- Information from Registrations;
- Personal and Corporate Income Taxes;
- Value Added Tax (VAT) and VAT Information Exchange System;
- Custom and Excise duties;
- Local government taxes and revenues;
- Electronic invoices data.

Mixed-sources models



- The **Split Data Approach**:
 - questionnaire only,
 - questionnaire supplemented by administrative register,
 - administrative data supplemented by questionnaire,
 - administrative data only
- **Pre-filled** questionnaires: respondents being asked to merely check and correct
- Using secondary sources for **non-responders** and for **estimation**



Integrated Metadata

- An integrated metadata system is a corner stone of a **multi-source** production system.
- Interconnecting metadata between internal production systems and processes is a must to produce integrated statistics.
- This is however a difficult harmonisation process.


Metadata System


<http://smi.ine.pt>


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 ...](#)

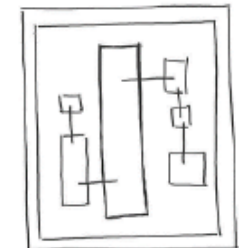
Code	Designation
10053	IUTICE - INQUÉRITO À UTILIZAÇÃO D...

Towards a data-centric perspective



Production chain: changing perspective

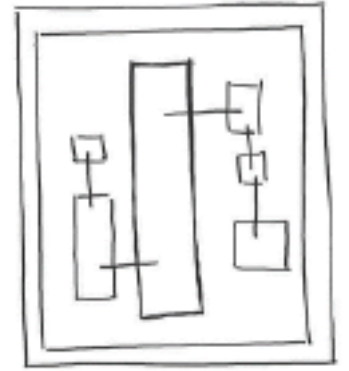
- Statistical organizations seek to produce high quality data that is supported by the metadata needed to make this data useful.
- It is often assumed that statistical production follows a certain chain of processes.
- It begins by collecting data, followed by processing, analysis and dissemination of statistical products.
- This is an inefficient approach to the present day as it consumes a lot of time and resources.
- Another perspective is needed.



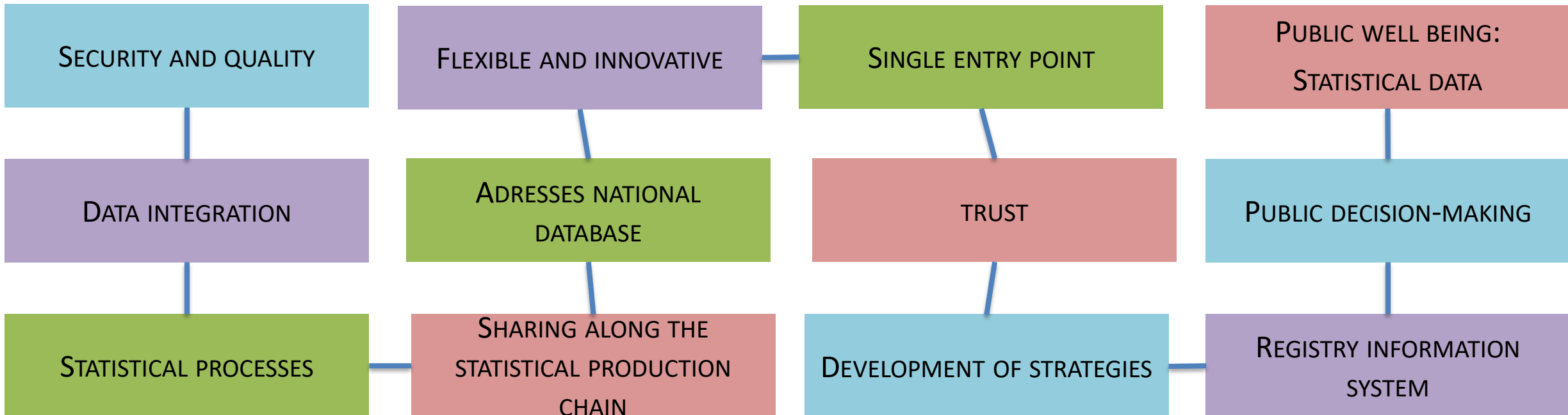
Data-centric production

- We can look at statistical production from a data-centric point of view;
- Focus on the **flow and the transformation of the data** during the production process;
- **Data Resources** should be understood as an **organizational asset**, which must be managed and explored intensively;
- **Reusing data** is certainly a way to make statistical production more efficient, both for existing products and for responding to new needs, at a lower cost and much faster.
- **Extending the use of mixed-sources** is a way to maximize de reuse of data.

National Data Infrastructure: a data-centric approach



MAIN GOALS



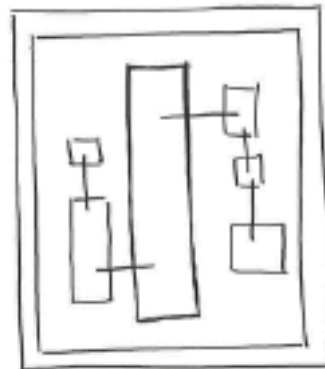
National Data Infrastructure: Challenges

MORE INTENSIVE AND INTEGRATED USE OF DATA

GAIN RESOURCES AND TIME TO INTENSIFY INNOVATION

SUBSTANCIAL ENLARGEMENT OF THE DOMAINS COVERED

LEARN THE MULTIDIMENSIONALITY OF DATA



NEW TECHNIQUES OF DATA EXTRACTION AND ANALYSIS

TO PROFIT FROM THE PRODUCTION CHAIN

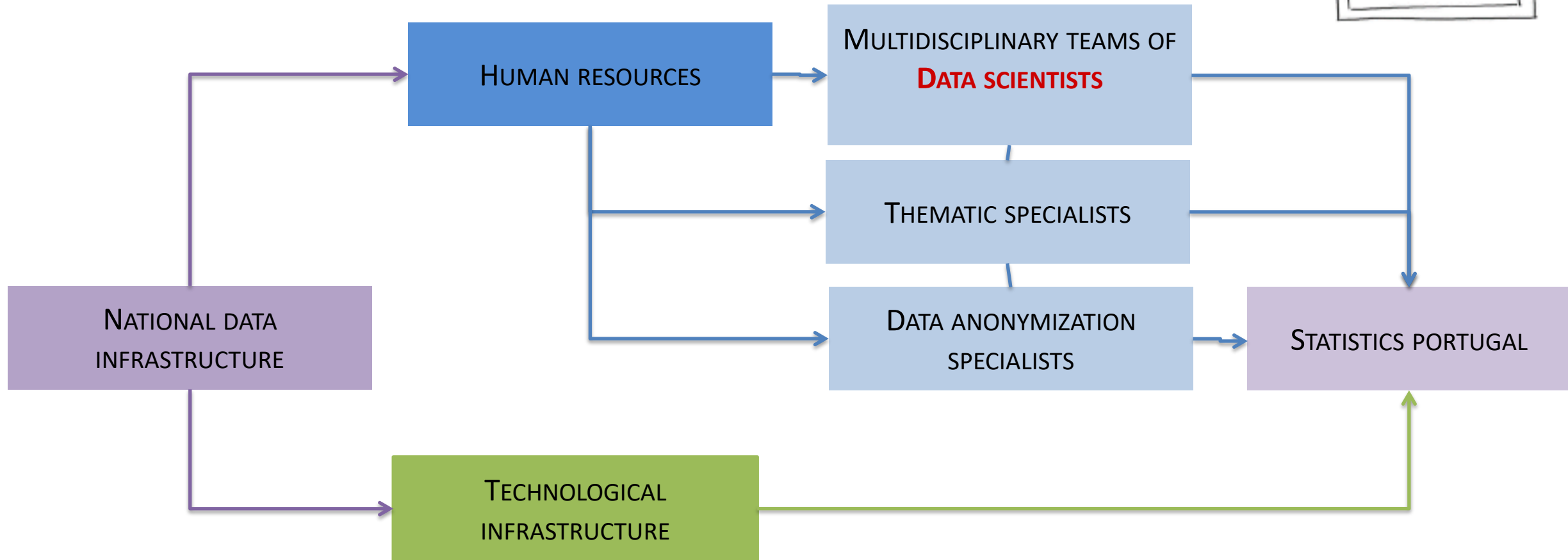
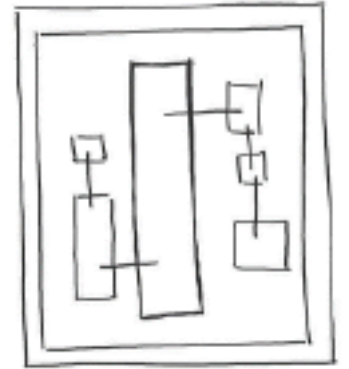
INTENSIFICATION OF APPROPRIATION OF ADMINISTRATIVE DATA AND OTHER SOURCES

INCREASE FEEDBACK TO SOCIETY



National Data Infrastructure

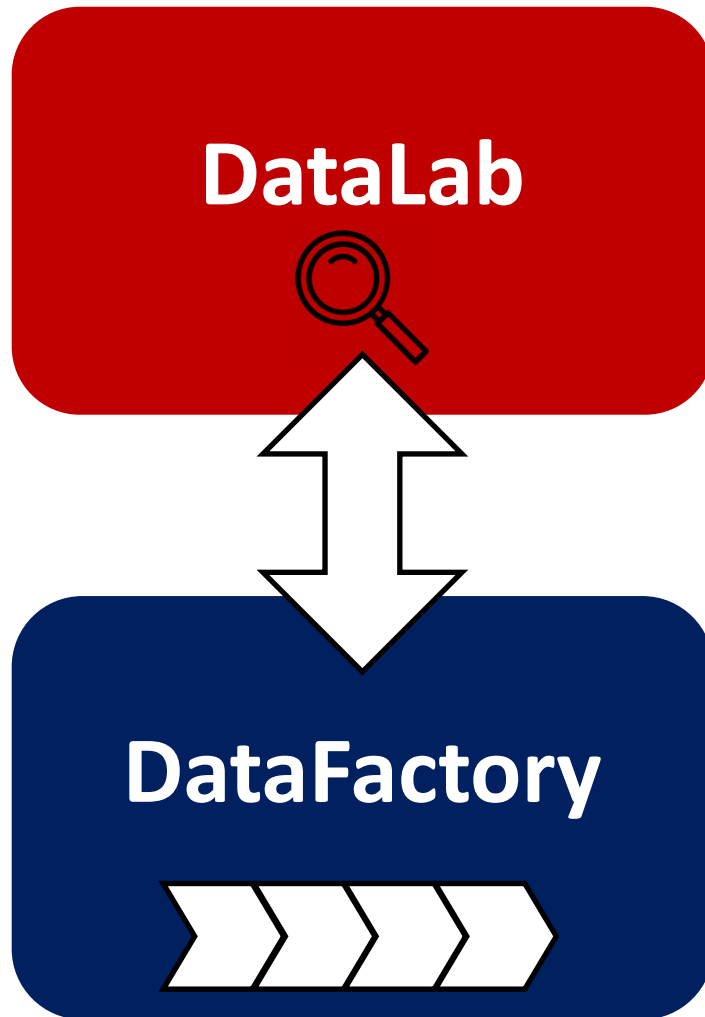
TECHNOLOGICAL INFRASTRUCTURE AND HUMAN RESOURCES



New perspective: DataLab and DataFactory



New perspective: DataLab and DataFactory



DataLab

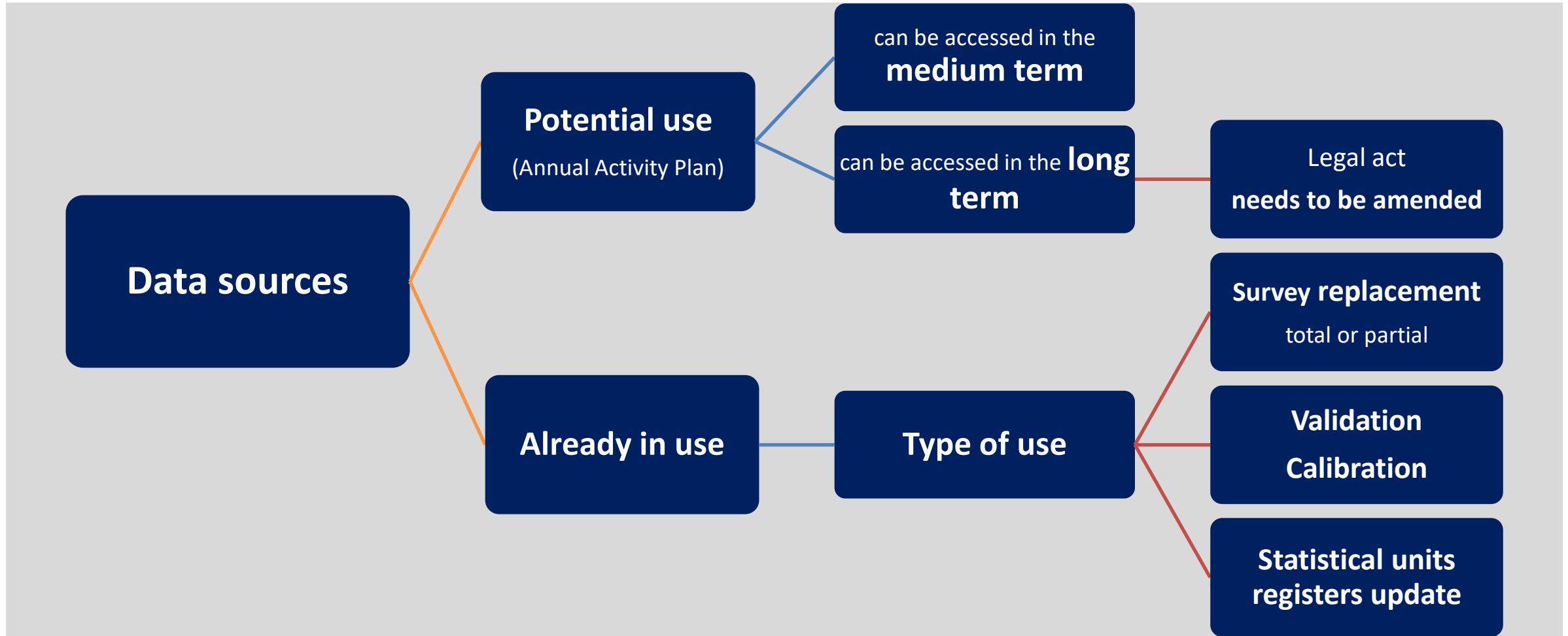
**Potential
Sources**

- Data Exploration
- Data Knowledge
- Specify Needs
- Check data availability
- Data Catalogue

DataLab



Data sources with **potential use** (or already in use)



Data access and data reuse barriers

Statistical Units

Access Stability

Variables

Classifications

Timeless

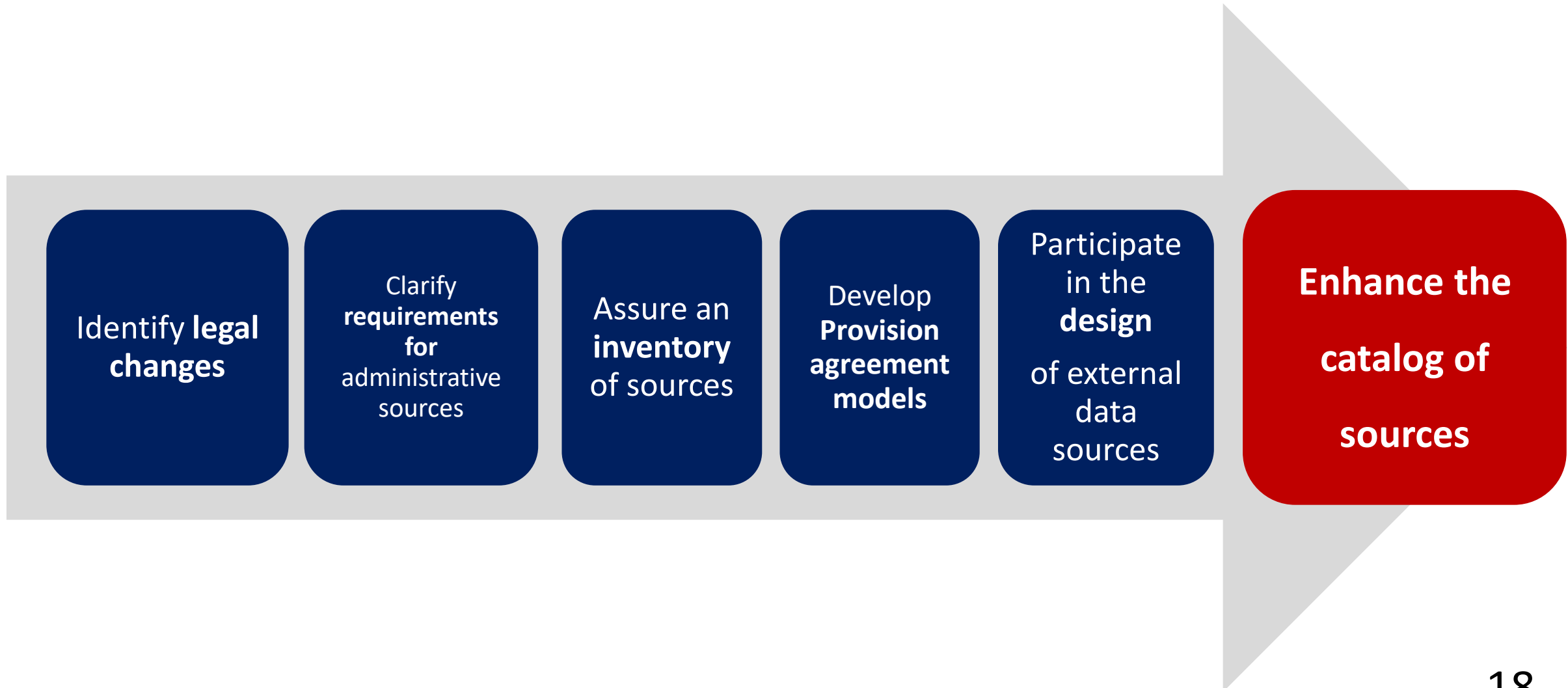
**Inconsistencies
between sources**

Missing Data

**Internal
Resistance**

Concepts

Administrative data appropriation process



Data Catalogue

- Checking **current data sources** could meet user requirements, and the conditions to be available, including any restrictions on their use.
- An **assessment of possible alternatives**, including research into potential administrative or other non-statistical data sources, to determine whether they would be suitable for use for statistical purposes.
- When existing sources have been assessed, a strategy for filling any remaining gaps in the data requirement is prepared.
- Should include an assessment of the **legal framework** in which data would be collected and used, and identify proposals for changes to existing legislation or the introduction of a new legal framework.

Challenge: Administrative Data Catalogue

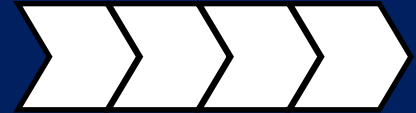
Source	Data Set	Flow	Variable	Agents	Existing Use	Potential
Provider	Data Description	Provision Agreement	Variable Definition	Activity Plan	Frequency	Timeliness
Technical Framework	Quality Indicators	Physical Location	Concepts	Agent Role	Historical Coverage	Transformation Level
Owner	Exchange Channel	Legal Framework	Statistical Units	Access Rules	Missing Data	Transformation Rules
Government Tutelage	Costs	Confidentiality Rules	Classification System	Access Tracking	Time Instances	Conversion Rules
Depositary Institution		Contingences				

DataFactory

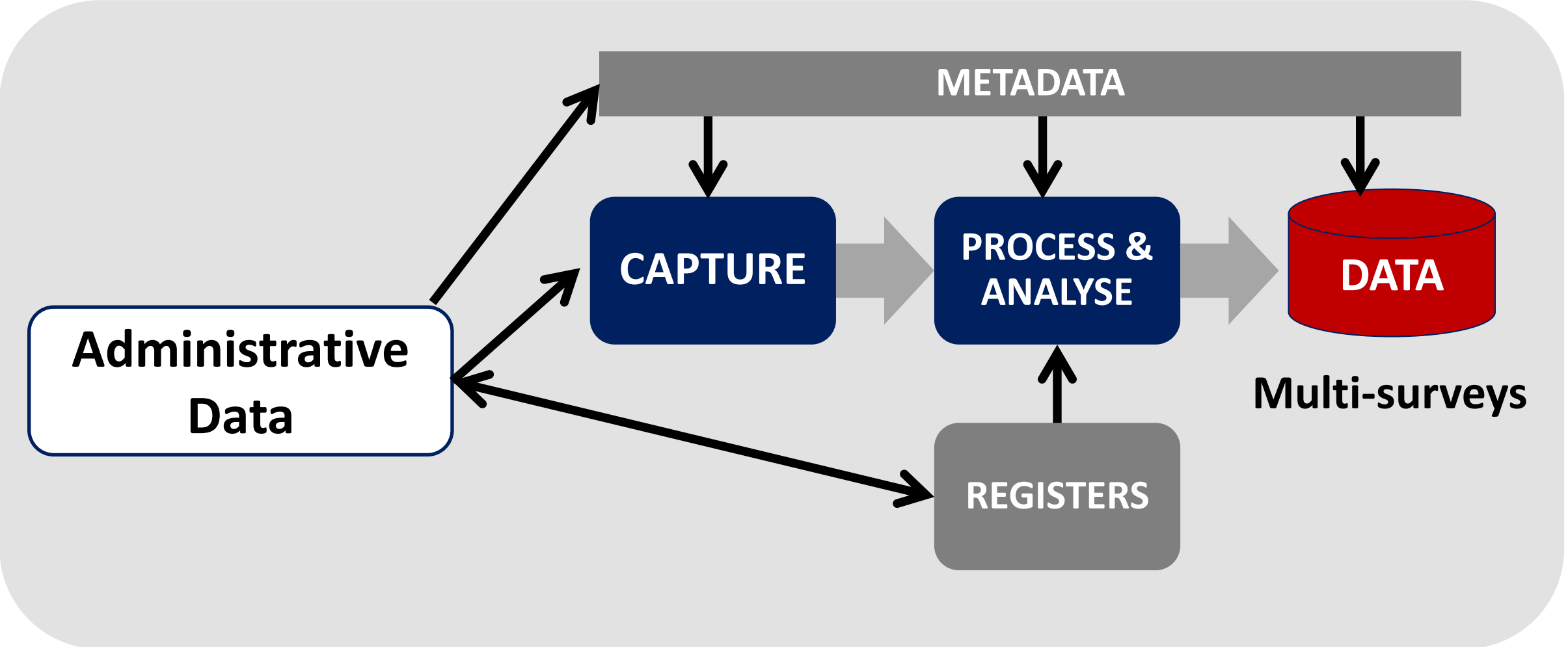
**Existing
Sources**

Data Processes

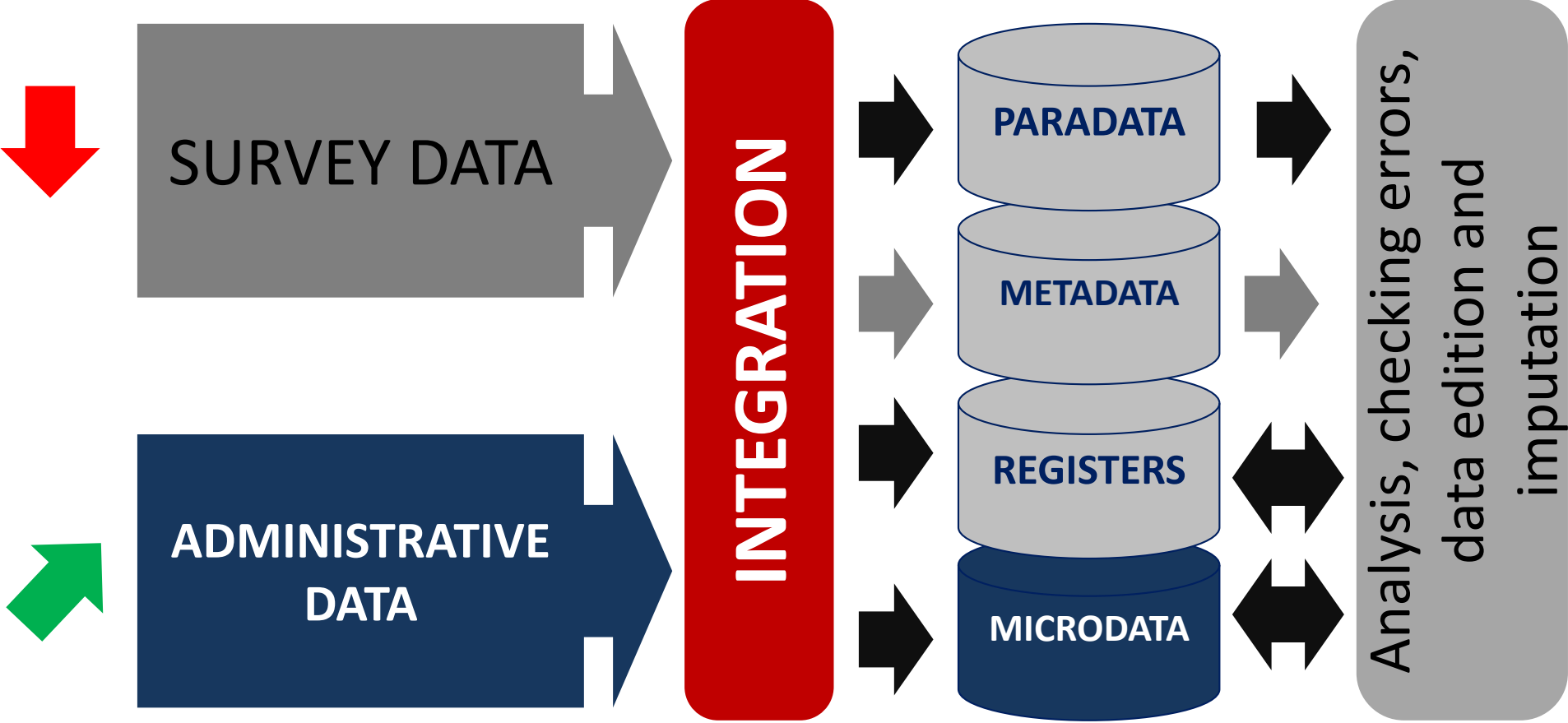
DataFactory



Administrative Sources – Processes



Source Integration





UNECE

modernstats
#HLG - MOS

Workshop on Statistical Data Collection 2019
14-16 October 2019, Geneva



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Multiple Sources in Business Surveys

thank you!



Paulo Saraiva

paulo.saraiva@ine.pt

Almiro Moreira

almiro.moreira@ine.pt

Sofia Rodrigues

sofia.rodrigues@ine.pt