# Trends of Data Collection processes in Official Statistics:

## *Specialization, Automation, Efficiency*

*Prof. Monica Pratesi*
*Istat*
*Director of the Department for statistical production*

2022 UNECE EXPERT MEETING ON STATISTICAL DATA COLLECTION "TOWARDS A NEW NORMAL?"

Rome, 26-28 ottobre 2022

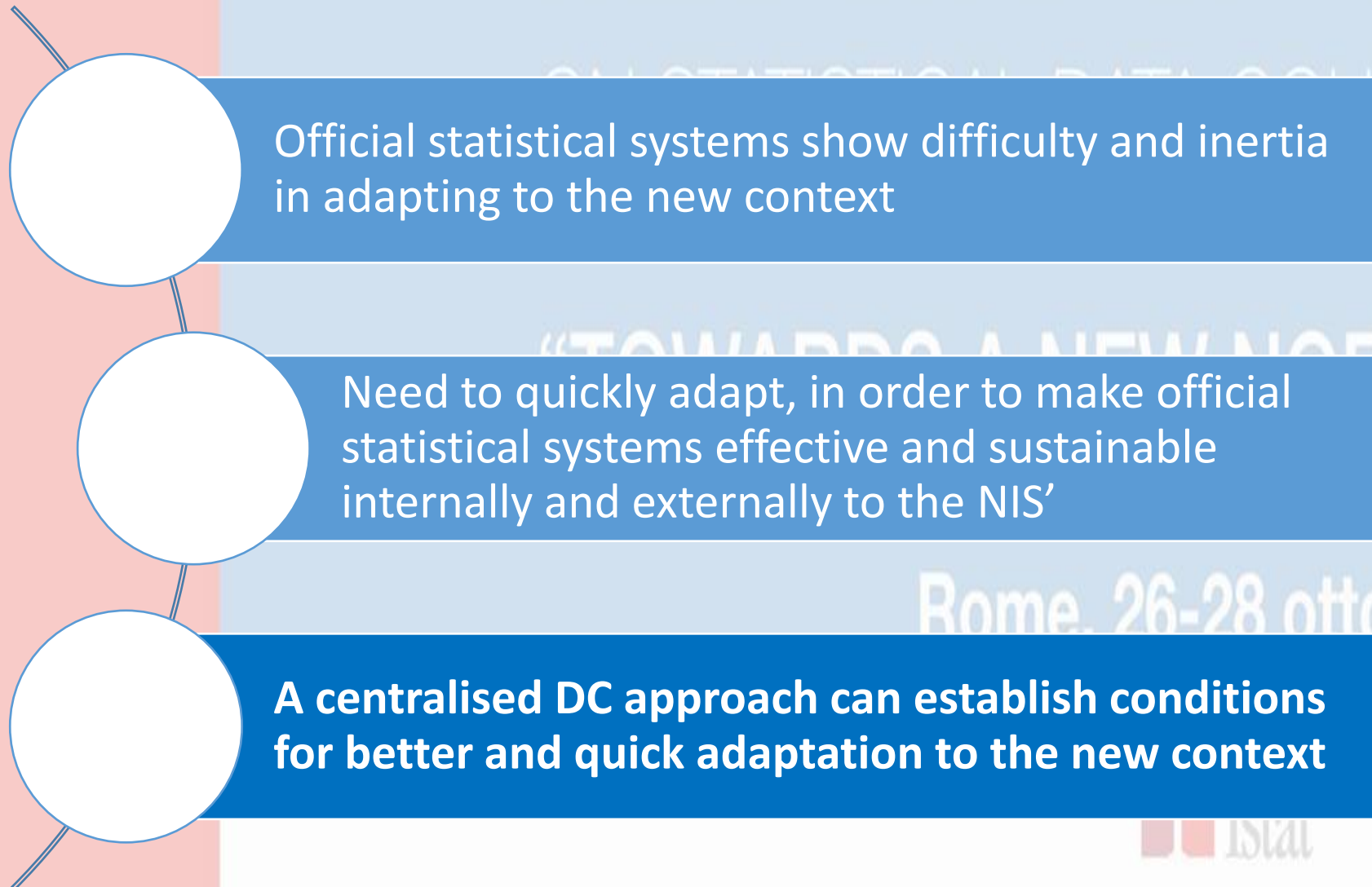# The framework of Official Statistics Data Collection (DC): current trends

## Internal trends

- **Increasing financial constrains and reduction of human resources available**

- **Strict quality requirements are confirmed**

- **More skilled and trained human resources**

## External trends

- **Strong technological and methodological innovative impulse**

- **Avalilability of new statistical sources**

- **Decreasing collaboration by the units involved in direct surveys> lower response rates**

# Adaptation of Official Statistics DC to the current trends

Official statistical systems show difficulty and inertia in adapting to the new context

Need to quickly adapt, in order to make official statistical systems effective and sustainable internally and externally to the NIS'

**A centralised DC approach can establish conditions for better and quick adaptation to the new context**

# The main features of the Centralised DC approach adopted by ISTAT

In order to face the current trends Istat implemented a centralised data collection model since 2016

For direct surveys this approach is opposed to the distributed one in which each survey adopted a specific and autonomous DC model, involving no or partial integration among DC processes

The centralization introduced a specialised approach that separate:

a) non thematic issues that are managed by integrated criteria for all/large groups of surveys
b) thematic aspects managed by specialised teams

Another branches of specialization of DC concern the acquisition of administrative archives useful to realise Integrated Registers, and the exploiting of new statistical sources

The aim of the new approach is to boost innovation and efficiency in Official statistics DC processes

The innovations involved technological, organizational and methodological issues

**The main dimensions of DC processes innovation**

Three dimensions of DC development in order to maintain tuning with the trends in progress

Specialisation

Automation

Efficiency

*The three dimensions are closely interrelated*

**DC Specialisation**

It concerns mainly the implementation of new methodologies

- The specialisation involves different thematic areas:
  - Multimode approach
  - Questionnaires design
  - Responsive survey design
  - TSE Total survey error framework
  - TSQ Total survey quality framework

*Relevant examples and case studies will be presented in sessions 1 and 3 of the UNECE DC2022 meeting*

# DC processes automation

**It is mainly related to technological developments**

- Several application areas can be identified:
  - Interoperability among registers and DB produced by different Institutions
  - Use of AI (Artificial Intelligence) techniques for the optimization of some data collection processes, e.g.:
    - Management of the centralized assistance service to the units involved in the surveys (Contact Center)
    - Application of "text mining" techniques
  - Design of APPs for DC purposes
  - Use of sensor data in order to replace or complement direct surveys
  - Implementation of generalised DC systems:
    - DC portals, integrated acquisition and survey management systems
  - Basic criteria: reuse and integration of tools and procedures

*Relevant examples and case studies will be presented in sessions 2 of the UNECE DC2022 meeting*

# DC efficiency

**It mainly concern the organizational assets of statistical processes**

- It involves analytic research work, oriented to standardization and integration, applied to the single phases of the DC processes e.g.:
  - Preparation of survey lists
  - Monitoring procedures
  - Formal and informal communications management
  - Penalties management
  - Design of a centralised Contact center for assistence
- Each of the above mentioned aspects is subject to generalization, standardization and optimization and requires a strong specialization of the resources used
- Efficiency improvements contributed substantially to the relevant increase of response rates of economic survey registered after centralization of DC in Istat (about + 11 percentage points for structural surveys ad about 20 percentage points for short-term surveys)

*Relevant examples and case studies are presented in sessions 2 of the UNECE DC2022*

# Relationship between centralisation and innovation

All dimensions and their related themes can be better planned, controlled and coordinated within a centralized DC model

Notably: costs, burden, quality

# Main strategic goals

Creation of networks among actors involved in Official Statistics DC

Maximising the exploitation of the potential offered by new technologies

Reducing the role of the direct survey in the production of Official Statistics by Increasing the use of alternative and innovative sources

Promoting the standardization and integration of DC processes in order to increase the efficiency and consistency of the information produced

Increasing the control over the different aspects of the TSE of direct surveys through the use of new methodologies

Optimization in the use of available resources and their skills

Reduction of the statistical burden on respondents to direct surveys

# The future

Trends in progress require timely adjustments by Official Statistical systems

In direct surveys the innovation of DC processes requires a clear and persistent impulse on the methodological, technological and organizational dimensions

The use of alternative sources and administrative archives has to be boosted in order to reduce whenever possible direct surveys

An integrated and coordinated approach to DC is necessary in order to optimize costs, statistical burden, quality of the information produced

A centralised DC model ensures exploitation of the opportunities offered by methodological and technological developments and the design of effective integrated and standardized solutions, as it involves a better governance on official statistical processes

Centralization is applicable on different scales: National statistical system, National Statistical Institute, thematic areas or homogeneous survey groupings: the main criteria is maximise re-use and efficiency of procedures

Istat    UNECE

# References

[1] Groves R.M. and Lyberg L. (2010), Total Survey Error: Past, Present, and Future Public Opinion Quarterly, Volume 74, Issue 5, 2010, Pages 849–879, https://doi.org/10.1093/poq/nfq065.

[2] Biemer P. Total survey error design, implementation, and evaluation Public Opinion Quarterly, Volume 74, Issue 5, 2010, Pages 849–879

[3] L. Rivais, M St-Denis, S. Lensen (2013), Centralising data collection at Statistics Canada. Seminar on Statistical data collection. Unece - Conference of european statisticians.

[4] P. Saraiva dos Santos, A.Moreira (2013), Creating a data collection department: statistics portugal's experience. Seminar on Statistical data collection, Unece - Conference of European statisticians.

[5] Snijkers, G., Haraldsen, G., Jones, J., & Willimack, D. (2013). *Designing and conducting business surveys*. John Wiley & Sons.

[6] Istat (2016), Istat's modernisation programme,https://www.istat.it/it/files//2011/04/IstatsModernistionProgramme_EN.pdf

[7]  Bavdaž, M., Snijkers, G., Sakshaug, J. W., Brand, T., Haraldsen, G., Kurban, B., ... & Willimack, D. K. (2020). Business data collection methodology: Current state and future outlook. *Statistical Journal of the IAOS*, *36*(3), 741-756.

[8] Bellini G., Bianchi G., Di Paolo GG., Papa P. *Towards a selective automation process of assistance to the survey units included in business surveys*, (2022) Business Data Collection Methods Workshop, Session T1 Language processing and machine learning techniques, Statistics Norway, Oslo, 13 - 15 June 2022.