

# 55.8 Utilizzo della fonte MDE nell'editing selettivo per le analisi delle asimmetrie



**30**  
NOVEMBRE  
2023

Francesco Amato, Francesco Ortame,  
Giuseppina Ruocco, Simona Toti

*#IstatWebinar*

# Indice della presentazione

---

- **Il contesto** – Giuseppina Ruocco
  - ✓ Obiettivi
  - ✓ Requisiti generali
  - ✓ Metodi
  - ✓ Metadati
  
- **Soluzione implementata** – Francesco Ortame
  - ✓ Illustrazione del software
  - ✓ Conclusioni

# Il contesto

Lo scambio di microdati relativi alle esportazioni intra-UE tra gli Stati membri (Micro-Data Exchange, MDE) offre agli Istituti Nazionali di Statistica una nuova fonte di dati per la compilazione e il benchmark delle importazioni intra-UE



# Obiettivi

- Approccio **standardizzato** e **armonizzato** all'editing dei dati nell'ambito del Sistema statistico europeo:
  - ✓ Utilizzo dei microdati MDE per implementare un nuovo sistema di produzione statistica armonizzato, al fine di individuare e riconciliare le **asimmetrie** bilaterali Import/Export
  - ✓ Passare a un sistema di produzione basato sui dati MDE richiede un'**analisi delle discrepanze** tra la fonte MDE (dati "specchio") e i dati delle importazioni raccolti a livello nazionale



- Implementazione di un **tool** armonizzato a livello europeo per individuare e riconciliare le asimmetrie statistiche intra-UE



# Requisiti generali

---

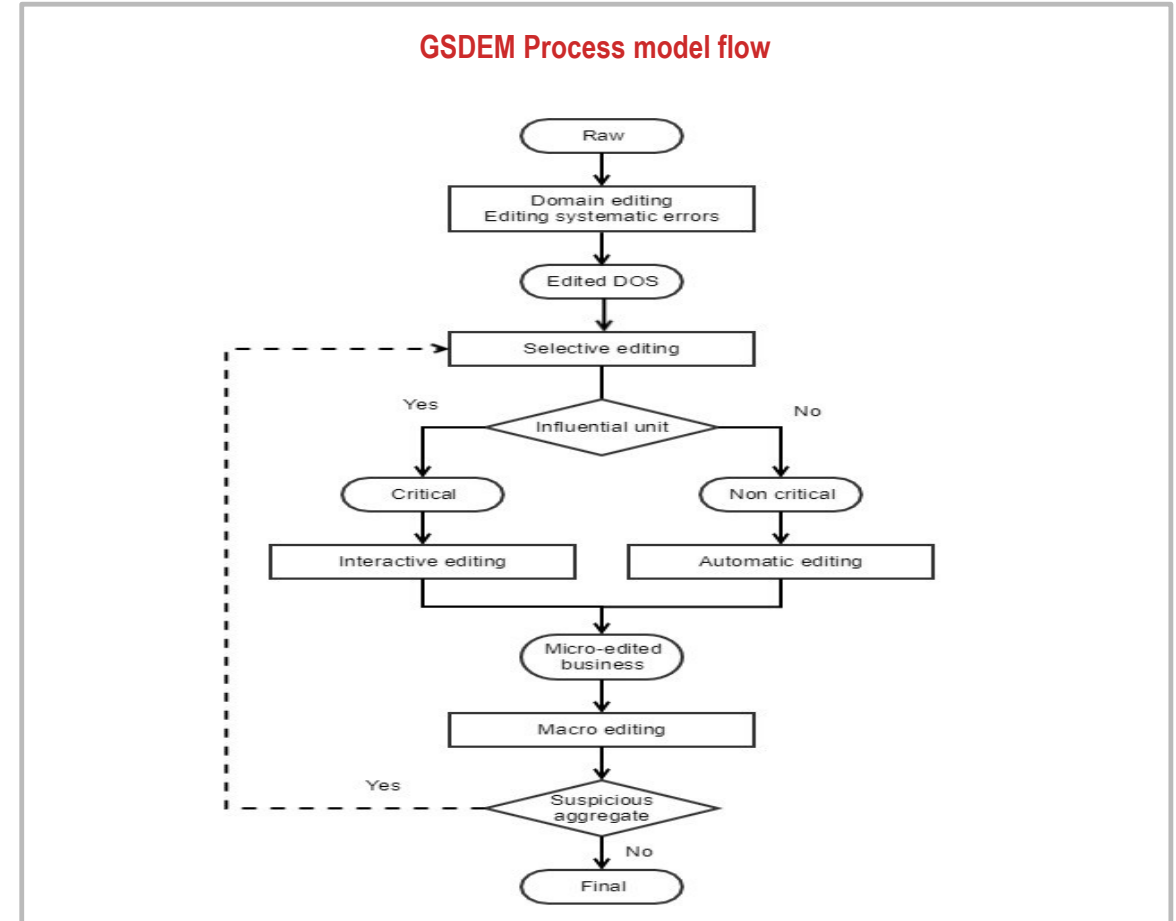
**Sviluppare uno strumento per l'individuazione delle principali discrepanze, sia per i prodotti che per i Paesi Membri**

- ✓ Implementare metodi condivisi per l'analisi delle asimmetrie mediante un **codice open source**
- ✓ Evidenziare le **unità influenti** e gli **errori sistematici**
- ✓ Facilitare la **riconciliazione delle asimmetrie rilevate**
- ✓ Riutilizzare l'analisi MDE, basata su un approccio top-down eseguito automaticamente, scomponendo l'asimmetria totale in termini di Paese, prodotto e operatore
- ✓ Costruire una **tassonomia delle potenziali asimmetrie**, per identificare per ogni tipo di errore e il metodo di validazione più appropriato

# Metodi (1)

A partire dallo standard della statistica ufficiale **GSDEM** (Generic Statistical Data Editing Model), i metodi implementati per l'individuazione delle asimmetrie sono:

- ✓ **Analisi esplorativa**, per studiare i modelli di distribuzione delle discrepanze tra le due fonti di dati
- ✓ **Rilevazione di errori sistematici**, per identificare gli operatori che, a parità di valori di Import, presentano codici di prodotto o di Paese diversi in MDE
- ✓ **Editing selettivo**, verifica finalizzata all'individuazione dei valori anomali, basata sul contributo relativo di ciascun valore all'asimmetria totale e su un indice per l'individuazione delle unità sospette



Fonte: [GSDEM](#) - Version 2.0, June 2019

# Metodi (2)

Nell'editing selettivo, il sottoinsieme di record da sottoporre a revisione interattiva è individuato mediante una soglia applicata a:

- ✓ **Contributo relativo di ciascun valore all'asimmetria totale**, calcolato come percentuale di  $D_i$  (differenza tra i dati nazionali e i valori MDE) sul totale assoluto  $D$ :

$$contr_i = 100 * \frac{D_i}{\sum_i |D_i|}$$

- ✓ **Indicatore di unità sospette**, calcolato all'interno di gruppi omogenei, identificati per Paese e classificazione dei prodotti a quattro cifre. Il primo e il terzo quartile ( $Q_1$  e  $Q_3$ ) sono calcolati sulla distribuzione empirica della differenza tra i dati nazionali e i valori MDE su scala logaritmica,  $\log R$  (Jader & Norberg, 2005)

$$S_i = \begin{cases} \frac{Q_1 - \log R_i}{Q_3 - Q_1} & \text{Se } \log R_i < Q_1 \\ \frac{\log R_i - Q_3}{Q_3 - Q_1} & \text{Se } Q_3 < \log R_i \\ 0 & \text{Altrimenti} \end{cases}$$

# Metadati

## Soluzione implementata e gestione dei metadati:

- **Metadati strutturali**, per standardizzare le strutture dei dati di input e preparare le variabili per l'esecuzione del codice
- **Metadati di processo**:
  - ✓ Soglia scelta per il cut-off dei record da revisionare manualmente
  - ✓ Numero di record risultanti dall'editing selettivo ed estratti per la riconciliazione delle asimmetrie
  - ✓ Percentuale di record da revisionare sul totale
- **Qualità**, da sviluppare a livello nazionale, per valutare l'accuratezza dei metodi di editing selettivo **dopo la fase di riconciliazione**

## Metadati e Qualità

### Metadati Strutturali

- Modello dati standardizzati

### Metadati di processo

- Soglia per il Cut-off delle unità da revisionare
- Record da revisionare in modo interattivo
- % di record da revisionare

### Qualità

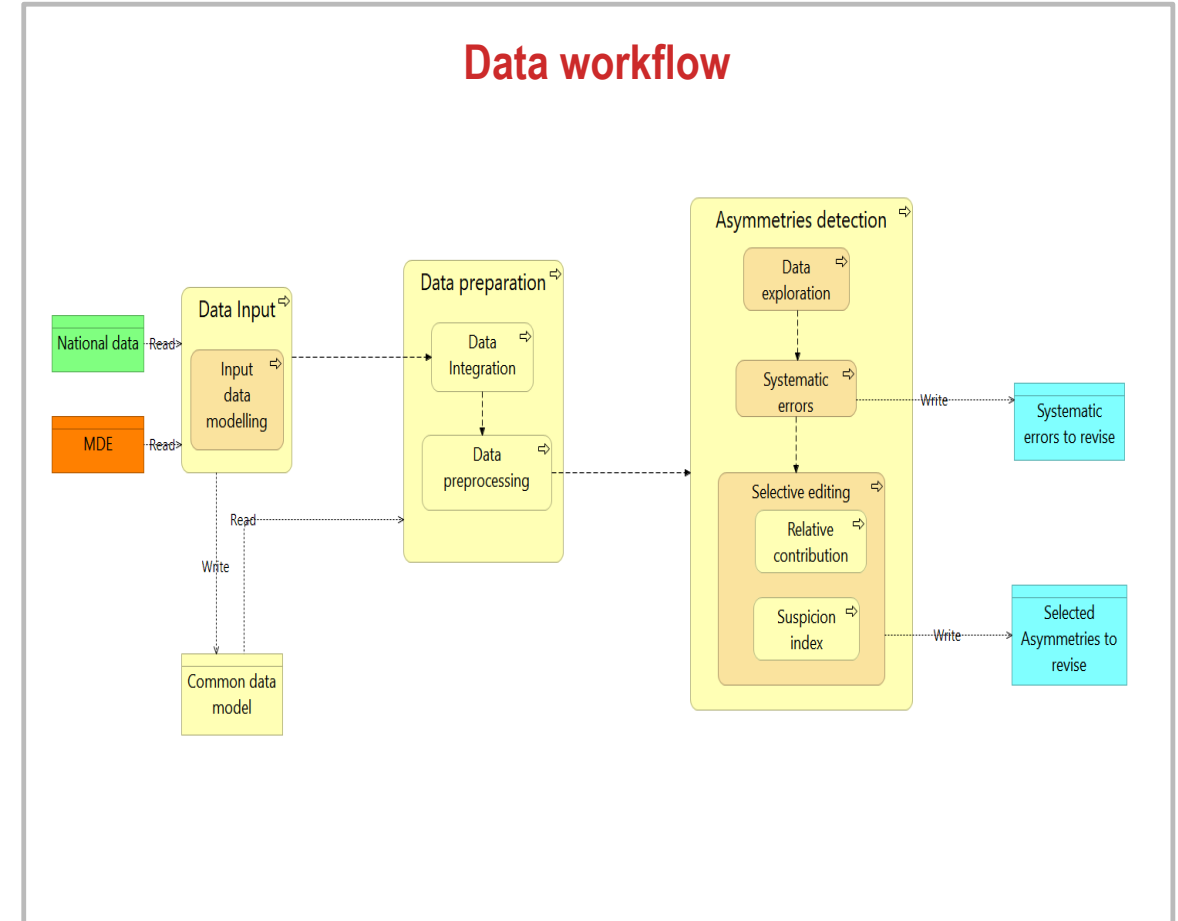
- % di errori effettivi sul numero di record risultanti dall'editing selettivo
- Confronto tra l'output del servizio e i risultati di procedure alternative per il rilevamento delle asimmetrie



# Soluzione implementata

## Step principali della procedura d'analisi implementata in R per l'individuazione delle asimmetrie

- ✓ **Input dei dati:** caricamento dei dati da elaborare e individuazione delle fonti secondo le strutture dei dati di input predefinite.
- ✓ **Pretrattamento dei dati:** mapping delle variabili, creazione di variabili aggregate, calcolo di indici descrittivi.
- ✓ **Individuazione delle asimmetrie:** analisi esplorativa dei dati, editing degli errori sistematici ed editing selettivo sulla base dei due approcci descritti. Dopo l'analisi, l'esperto di dominio può scaricare un file csv con i record problematici da controllare manualmente



# Conclusioni

---

- Il lavoro rappresenta un primo passo verso un sistema di produzione statistica basato su dati MDE per l'analisi delle asimmetrie
- La riconciliazione delle asimmetrie è difficilmente standardizzabile in una procedura, a causa della prevalenza di editing interattivo, che può differire a seconda del contesto nazionale e della disponibilità di fonti ausiliarie da utilizzare come benchmark, come il Registro delle Imprese
- Le soluzioni open source garantiscono trasparenza e agevolano la collaborazione tra statistici ed esperti di dominio, al fine di soddisfare esigenze specifiche degli utenti
- Questo approccio facilita il riuso delle soluzioni implementate e permette di affrontare le problematiche più comuni di interoperabilità, come l'integrazione dei servizi nella produzione statistica, la gestione delle dipendenze e del versioning, il supporto agli utenti

# Riferimenti bibliografici

---

- UNSTATS Methodology and Manuals: trade asymmetries
- Generic Statistical Data Editing Model (GSDEM - Version 2.0, June 2019)
- Common Statistical Data Architecture (CSDA Version 2.0)
- Generic Statistical Business Process Model (GSBPM) v. 5.1. January (2019)
- Hidiroglou M.A. e Berthelot J.M. (1986). Statistical Editing and Imputation for Periodic Business Surveys, Survey Methodology, Statistics Canada, June 1986, vol.12, N.1, pp.73-83.
- Jäder A., Norberg A. (2005), A selective editing method considering both suspicion and potential impact, developed and applied to the Swedish foreign trade statistics, paper presented at the Work Session on Statistical Data Editing, Ottawa, Canada, 16-18 May 2005.

# grazie

FRANCESCO ORTAME | [francesco.ortame@istat.it](mailto:francesco.ortame@istat.it)

GIUSEPPINA RUOCCO | [giuseppina.ruocco@istat.it](mailto:giuseppina.ruocco@istat.it)