CONFERENZA GARR 2024
Navigare la complessità

La piattaforma Big Data in Istat nei processi statistici e di sperimentazione

Marco Polizzi
Francesco Altarocca
Istat

L’Istat nella trasformazione digitale

La **missione** dell’Istat è quella di servire la collettività attraverso la produzione e la comunicazione di **informazioni statistiche**, **analisi** e **previsioni** di elevata qualità in ambito economico, sociale e ambientale

L’Istat opera nel pieno rispetto delle **normative** nazionali ed europee

L’Istat recepisce il **Piano Triennale** per l’informatica AgID

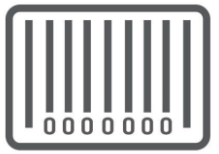
Sviluppo di **piattaforme tecnologiche abilitanti** all’utilizzo dei dati statistici e amministrativi

I processi di produzione statistica - Gli scanner data

Supporto IT ai processi di produzione statistica, tecnologie abilitanti alle **nuove fonti** di **dati**

Disponibilità di grandi quantità di dati, in particolare della **G**rande **D**istribuzione **O**rganizzata

- L'Istat aggiorna annualmente l'elenco dei prodotti che compongono il **paniere**
- Dal 2018 l'Istat utilizza i prezzi registrati alle casse della GDO mediante scannerizzazione dei **codici a barre**
- Rilevazione dell'**I**ndice dei **P**rezzi al **C**onsumo ⇨ **Inflazione**



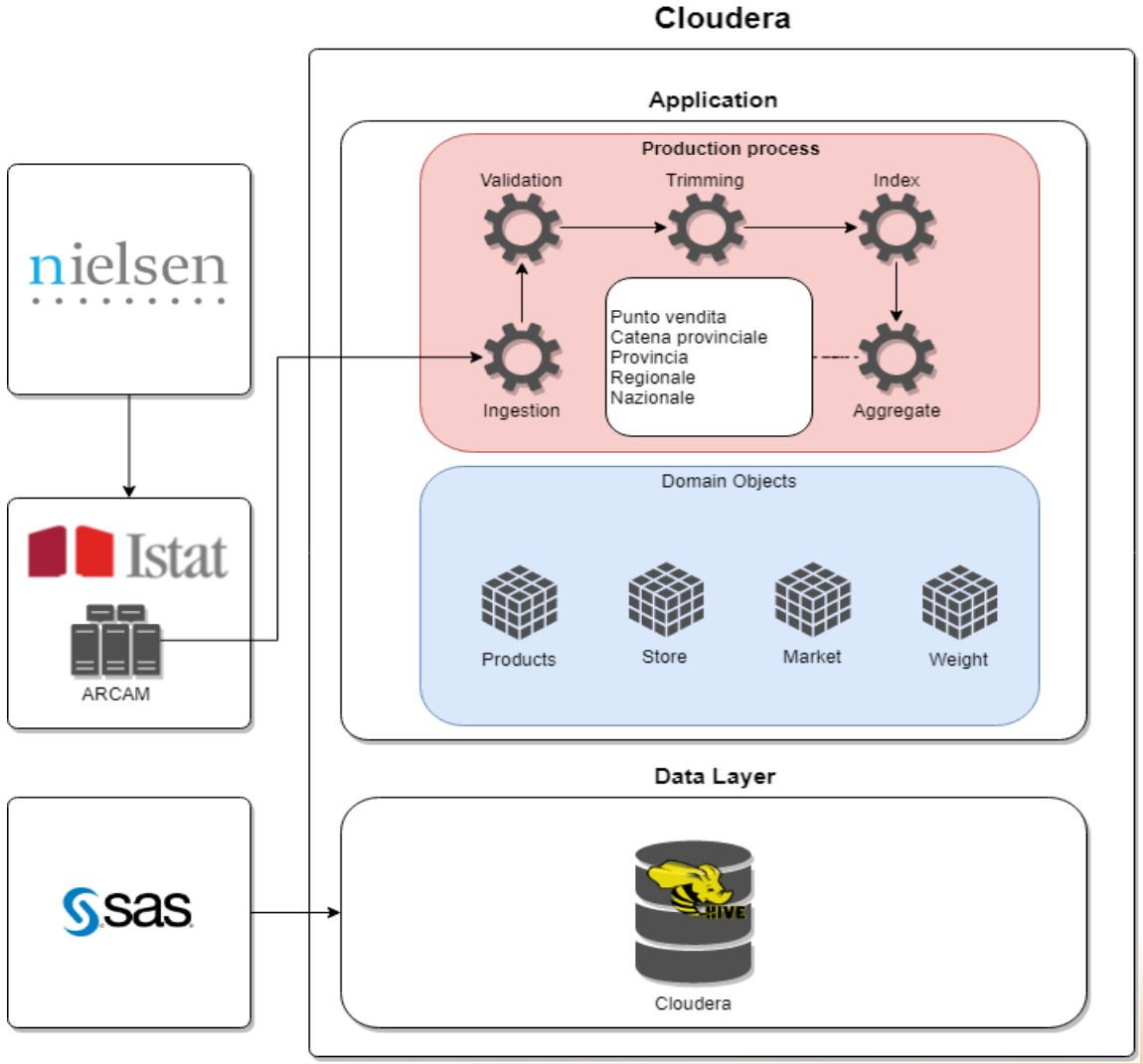
L’indice dei prezzi al consumo

4.000 punti vendita

- ✓ tutte le province

Nel 2023:

- ✓ 5 tipologie distributive della GDO
- ✓ 1 miliardo di dati
- ✓ 421 milioni di micro-indici
- ✓ 7 milioni di aggregazioni per livello territoriale
- ✓ 80 milioni di aggregati/anno



La stima della povertà e la parità regionale del potere di acquisto

Povertà

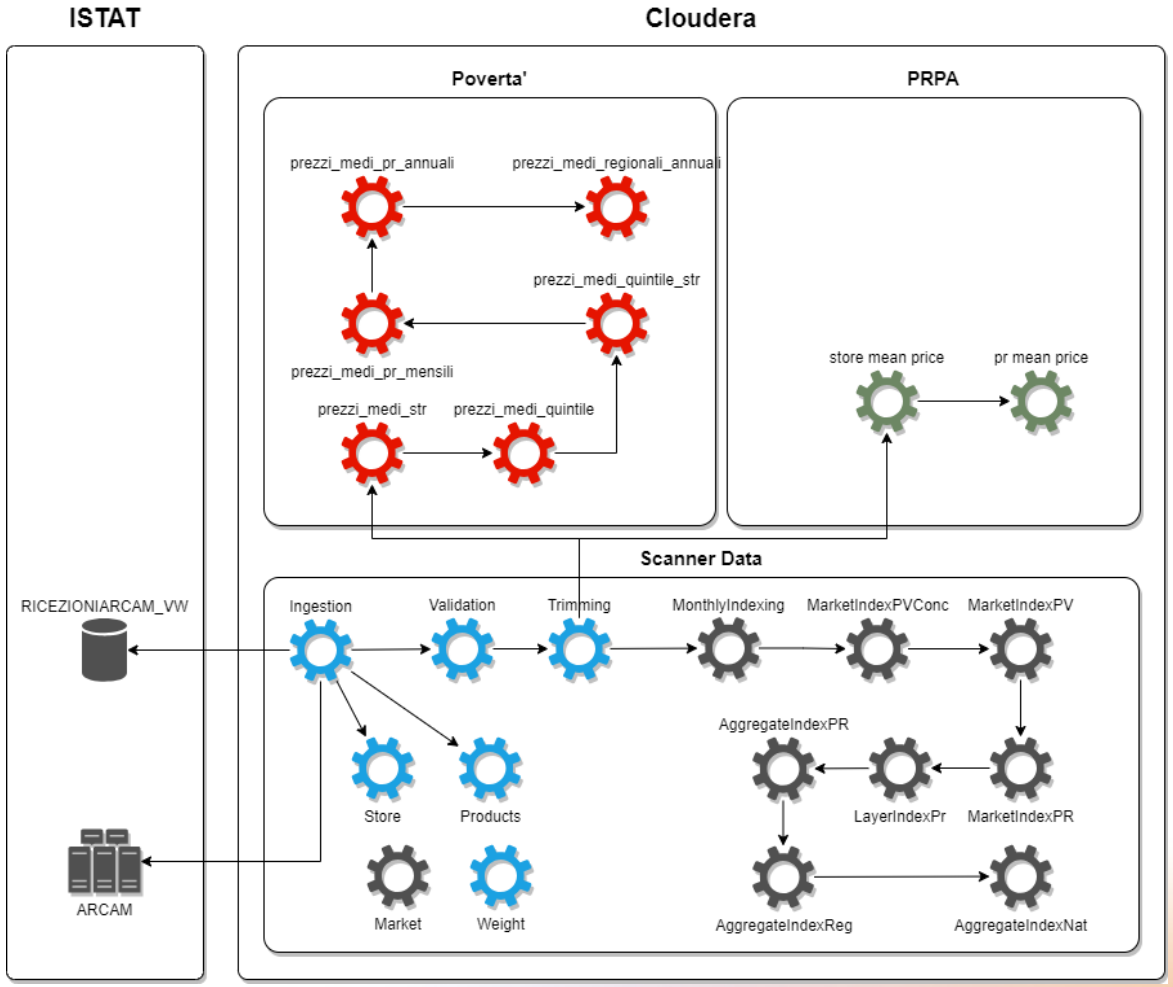
- Dati elementari già acquisiti per l'indagine sugli IPC con aggiunta del formato del prodotto

- Stima della povertà assoluta

Parità regionale del potere di acquisto

- Misura le differenze nel livello medio dei prezzi di un paniere di prodotti fra una regione e l'altra

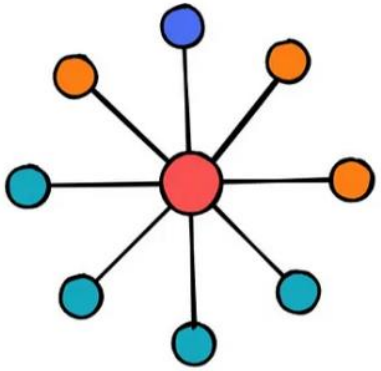
- Diseguaglianze e condizioni di vita delle famiglie

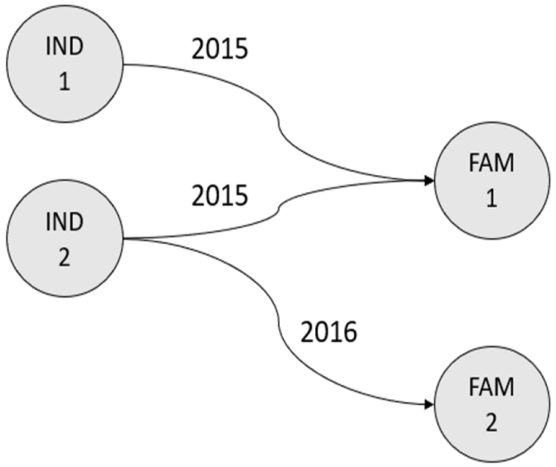


Le sperimentazioni per lo sviluppo di metodologie statistiche

Il record linkage probabilistico

- Test di performance (7 gg su 7, h 24) sul record linkage con la piattaforma big data
- Utile per lo studio di fattibilità di altri progetti per l'elaborazione di quantità di dati finora difficilmente gestibili





Le basi di dati a grafi

- Test sulle potenzialità dei DB a grafo (vs RDBMS) nel collegare entità e caratteristiche degli individui in layer differenti
- Utile per la facilità e l'efficacia di estrazione di struttura, relazioni e spostamenti nel tessuto delle famiglie

La piattaforma Big Data in Istat

- Definizione di strutture dati indipendenti
- Velocità ed efficacia nella gestione di grandi quantità di dati
- Riuso di pacchetti e procedure per altre indagini statistiche
- Integrazione con altri processi di produzione statistica e di ricerca

Istat - Direzione centrale per le tecnologie informatiche

Marco Polizzi - polizzi@istat.it

Francesco Altarocca - fraltaro@istat.it