



A Scuola di **OPENCOESIONE**

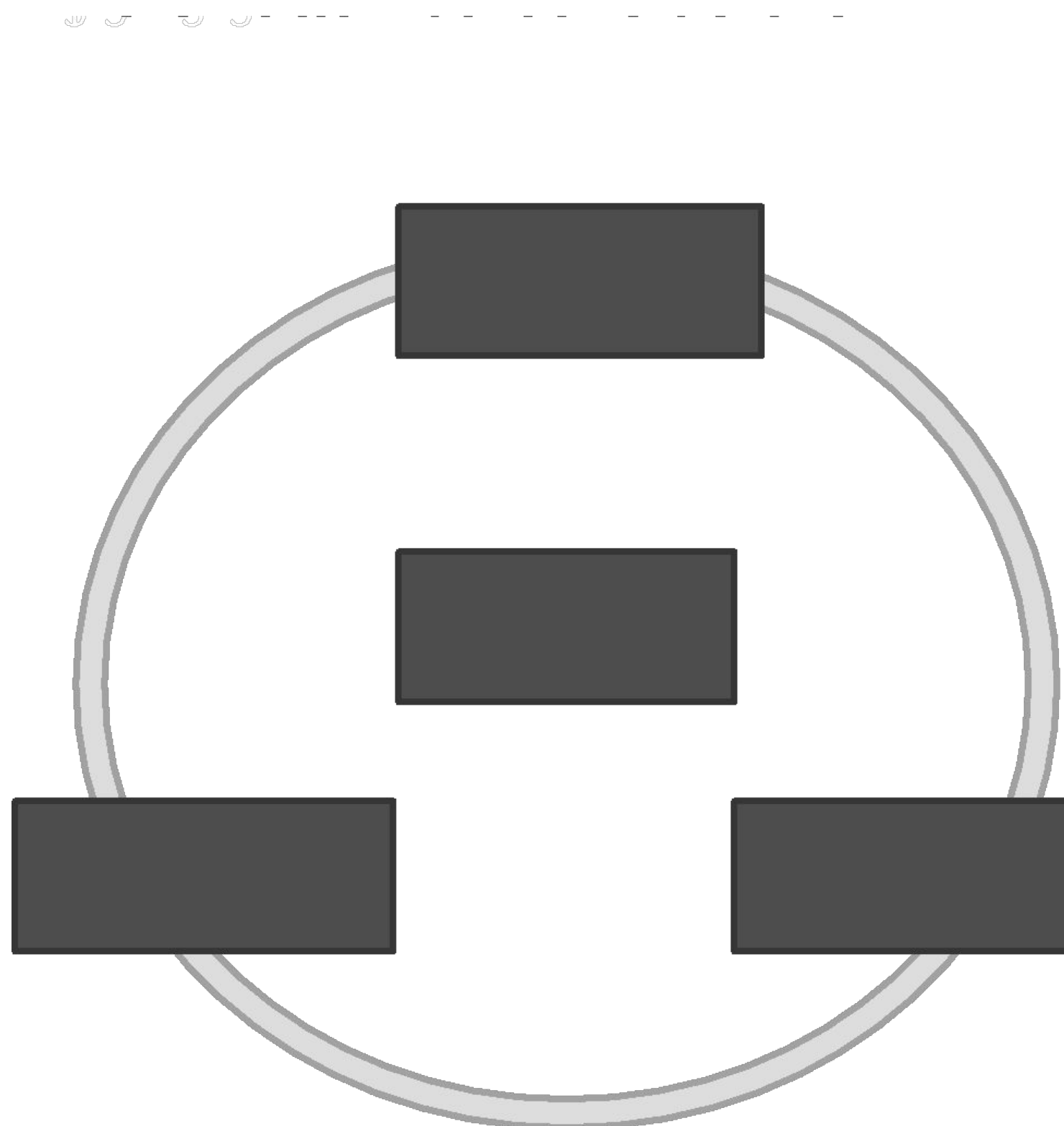
ISTATDATA: LA NUOVA PIATTAFORMA ISTAT PER LA DIFFUSIONE AGGREGATA

Carlo Boselli - Istat

30 novembre 2023

- La scelta di una piattaforma di diffusione
- Contesto e strategie nella diffusione dei dati aggregati Istat
- La soluzione tecnologica: Data Browser, Meta Data Manager

La scelta di una piattaforma di diffusione: Approccio multidisciplinare



Si interviene su un sistema complesso:

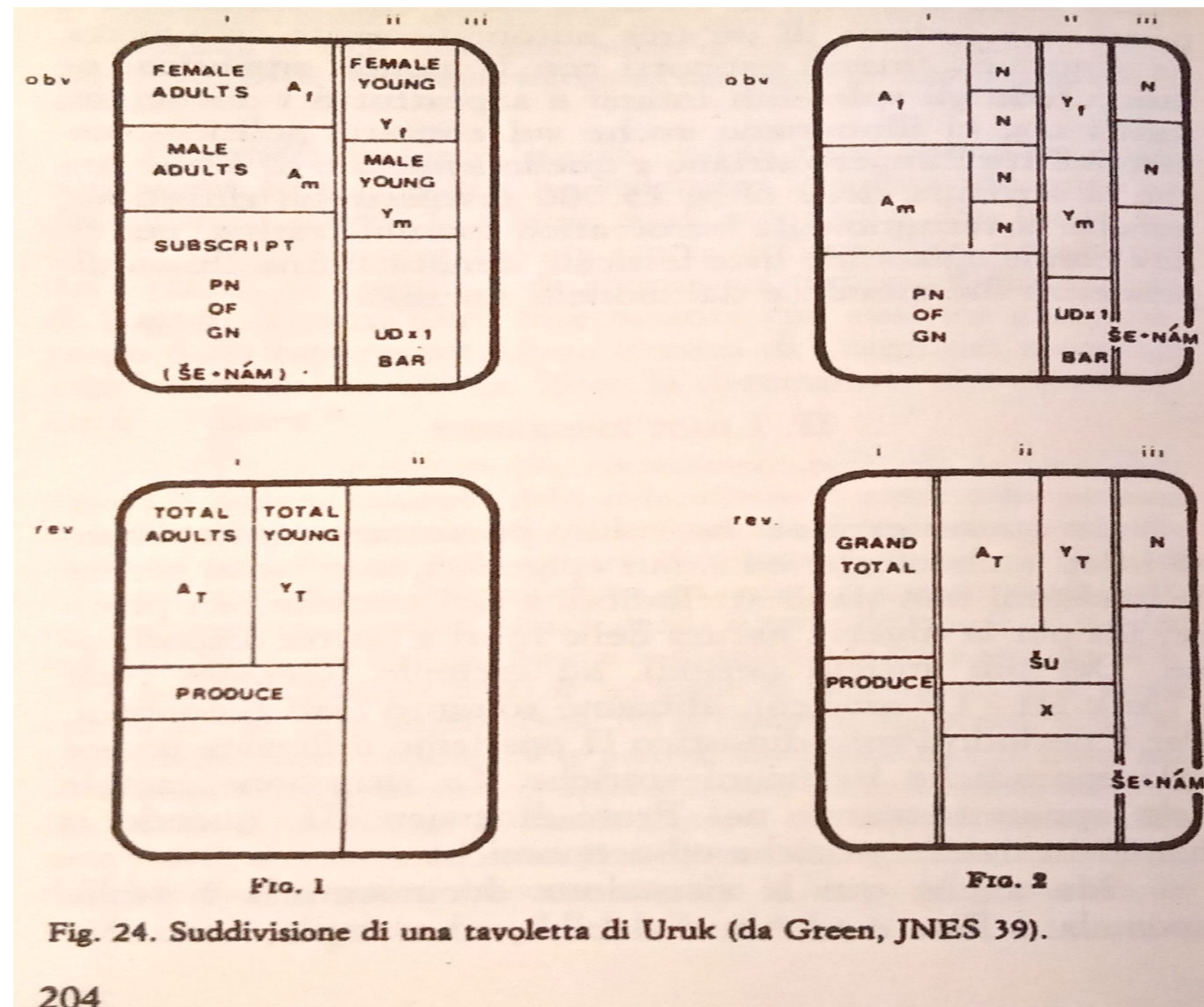
- Diversi tipi di dato: amministrativi, da indagine, dati micro e aggregati
- Diverse strategie e processi nella stessa organizzazione

In generale non esiste una tecnologia migliore per ogni forma di diffusione, ma quella che di volta in volta presenta il miglior bilanciamento tra diversi aspetti.

La tecnologia deve adattarsi agli obiettivi, alla caratteristica dei dati e dei processi.

Approccio multidisciplinare:

Esperti di organizzazione aziendale, statistici, modellatori di dati, esperti di informatica.



Struttura delle tavole per la diffusione dei dati agricoli - (M.W. Green, Journal of Near Eastern Studies 39 - Animal Husbandry at Uruk in the Archaic Period), in «I Sumeri» - Giovanni Pettinato - Bompiani

URUK III 4.500 anni fa, tavolette di argilla per la diffusione dei dati:

- Rappresentazione dei dati mediante una struttura estesa e ridotta;
- Suddivisione degli animali per sesso, per età, indicazione degli animali nati dai precedenti, la relativa produzione di burro e formaggio, i responsabili della registrazione e delle informazioni amministrative.

Alcuni aspetti della diffusione sono connaturati alle esigenze degli utenti, alla natura dei dati, al sistema organizzativo:

- Metadattazione, modellazione dei dati, confronto e collegamento dei dati
- Standardizzazione e stabilità del processo di diffusione

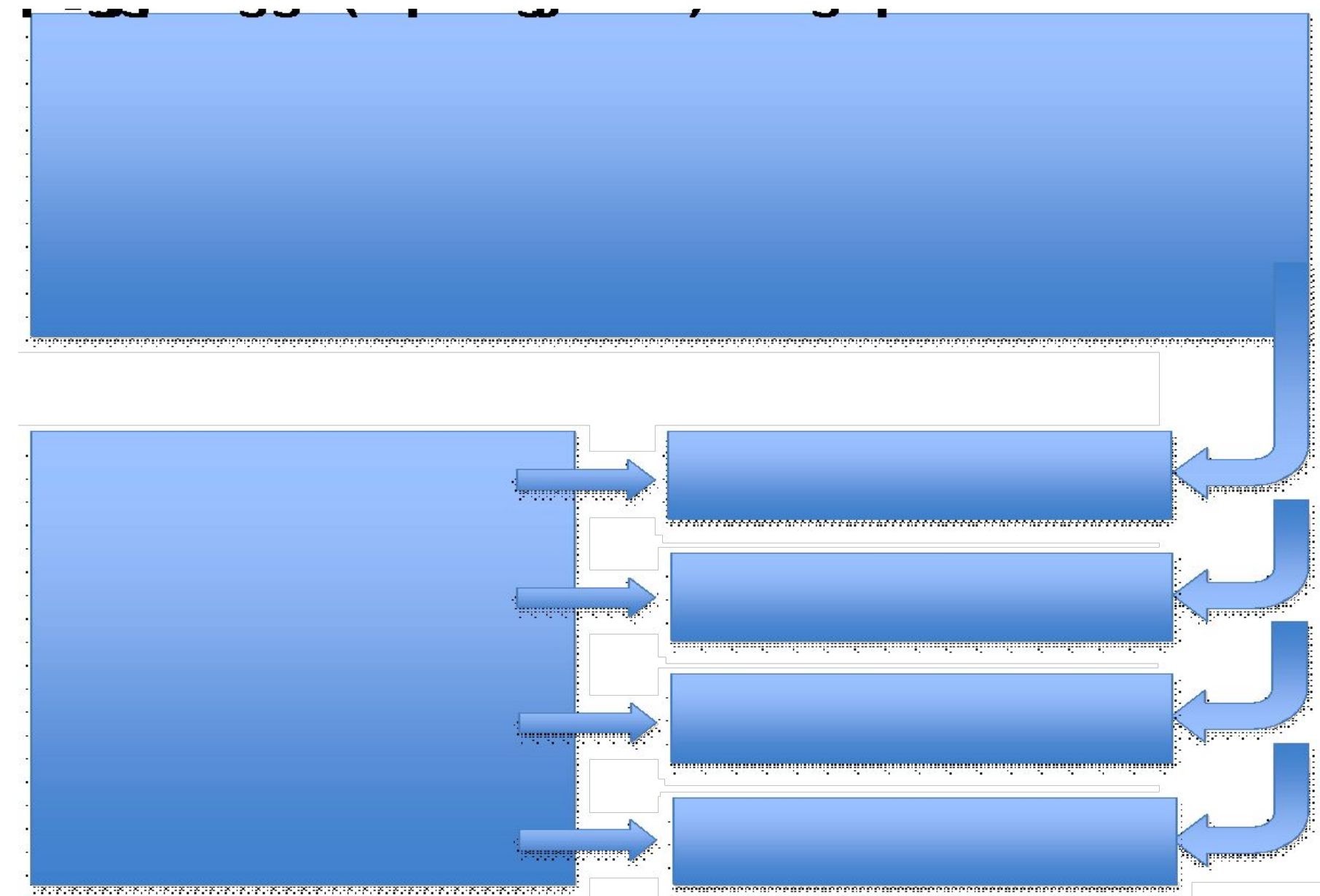
Considerare il contesto, le strategie e le soluzioni tecnologiche

Contesto organizzativo e strategie

Utenza esterna (Integrazione e granularità)



Eurostat (Indagini sotto regolamento)



Contesto organizzativo e strategie

Contesto specifico in Istat per la diffusione su larga scala dei dati aggregati:

- Dati aggregati provenienti da un gran numero di indagini sotto regolamento Eurostat i cui risultati devono essere tutti diffusi. In alcuni casi i dati aggregati derivano da una combinazione di dati amministrativi e di indagini.
- Un gran numero di processi già esistenti di aggregazione e diffusione dei dati
- La struttura di diffusione dei dati è pressoché stabile nel tempo
- Dati provenienti da indagini diverse che possono essere integrati solo parzialmente per specifici domini di analisi

Contesto organizzativo e strategie

Strategie:

1. Permettere agli utenti di avere un ambiente omogeneo e ridurre i costi di accesso
2. Garantire flessibilità nel confrontare dati aggregati pre-strutturati
3. Soluzioni di rappresentazione sintetica dei dati senza modificare la struttura dei dati e i processi sottostanti
4. Elevata capacità di archiviazione dei dati:
5. Migrare ad uno standard (SDMX) per:
 - sfruttare la modularità tecnologica senza modifiche nella modellazione dei dati e preservando i processi di diffusione dei dati in futuro;
 - consentire l'accesso «machine to machine» utilizzando l'API

La soluzione tecnologica: Data Browser, Meta & Data Manager

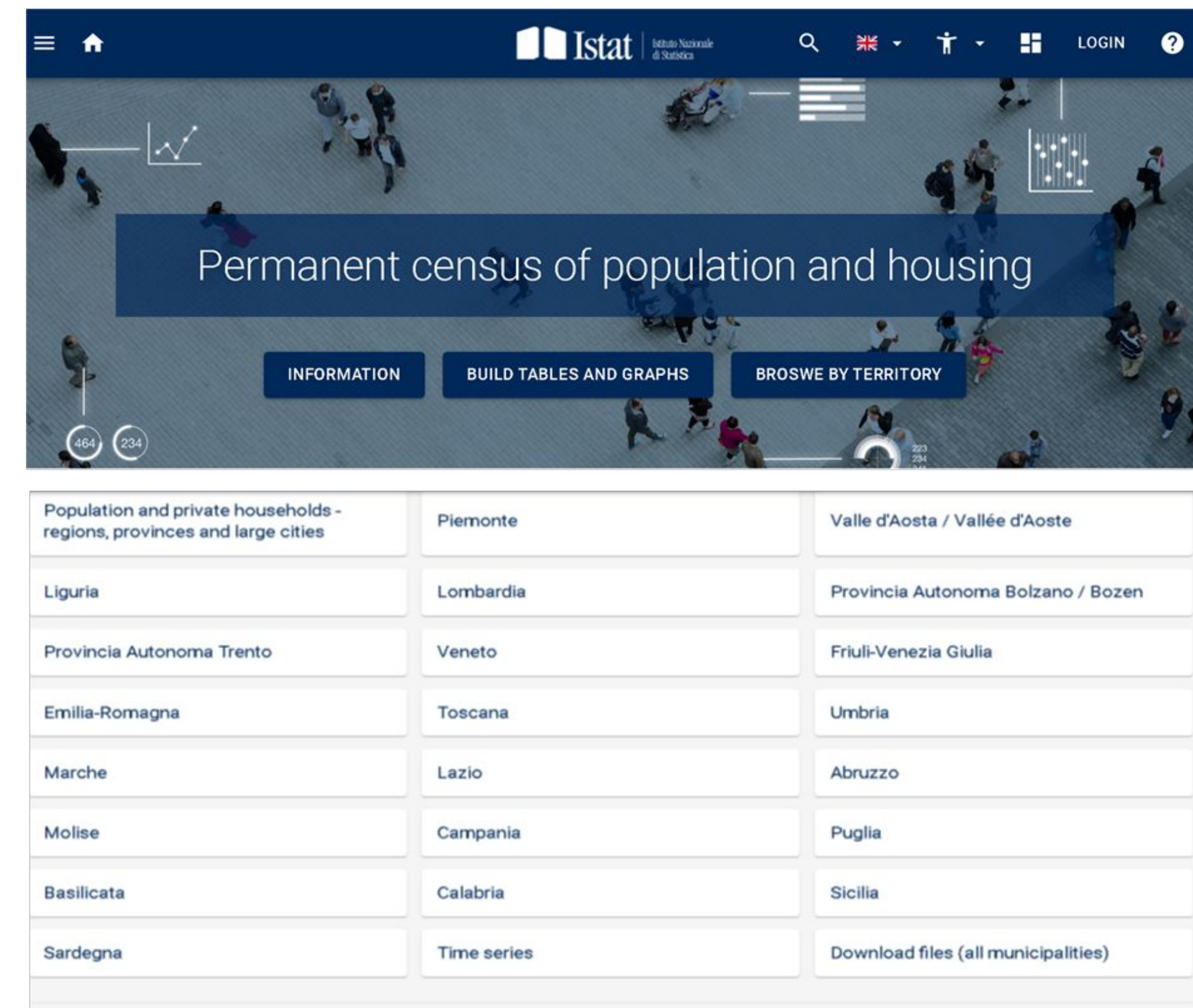
Piattaforma open source sviluppata dall'Istat e basata su due componenti:

1. **Data Browser** (Front end): per consentire agli utenti esterni di consultare i dati
2. **Meta & Data Manager** (Back end): per creare strutture dati SDMX, cubi dati, mappatura dati, caricamento dati e flussi dati (per utenti interni Istat)

Le tecnologie front-end e back-end si basano sullo standard SDMX.

A partire dal 2020 i dati del Censimento permanente della popolazione e delle abitazioni sono diffusi sulla nuova piattaforma:

<https://esploradati.censimentopopolazione.istat.it>

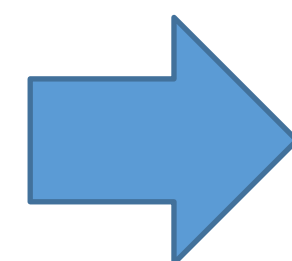


Sistemi di diffusione migrati verso la nuova piattaforma

M&DM and DATA BROWSER

Sistemi precedenti Istat per la diffusione dei dati aggregati

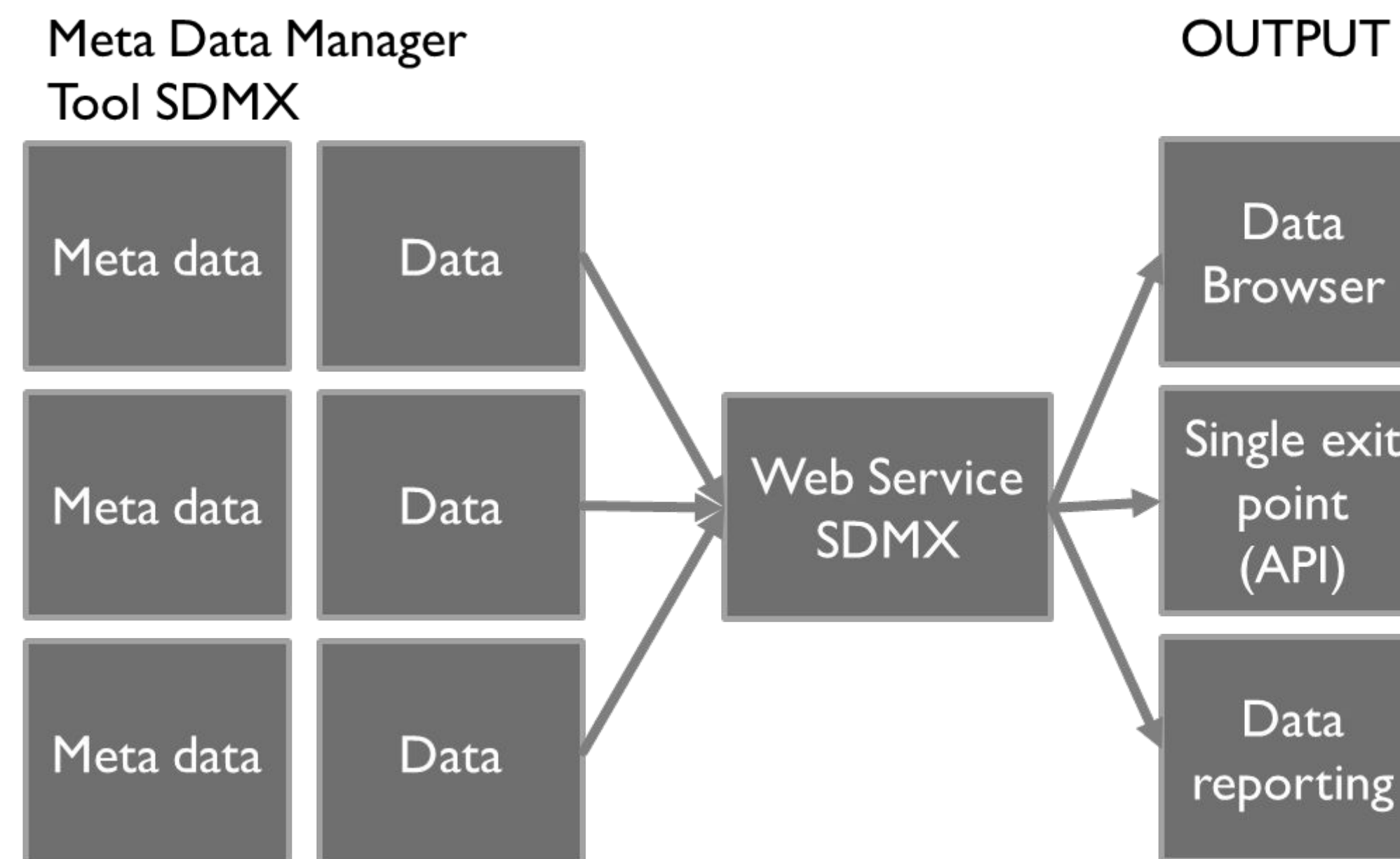
- I.Stat
- Censimenti permanenti
- Sistemi tematici (Giovani, Anziani, Capitale umano, Pubblica amministrazione ...)
- COEWEB (Commercio estero)



1. **IstatData** – La banca dati dell'Istituto nazionale di statistica (<https://esploradati.istat.it>). Più di 450 datacubes – **2 miliardi di record (punti dato)**
2. **Statistiche sul commercio estero (Beta version)** (<https://esploradati.istat.it/coeweb>) . Più di 47 cubi di dati – **10 miliardi di record (punti dato)**
3. **Censimento permanente della popolazione e delle abitazioni** (<https://esploradati.censimentopopolazione.istat.it>)

DATA MANAGER: webservice e single exit point

- Architettura basata nativamente su uno standard di trasferimento dati riconosciuto a livello internazionale (SDMX)
- Elevata capacità di archiviazione dati per singolo datacube
- Infrastruttura modulare con accesso tramite webservice
- Il nuovo sistema tecnologico è sviluppato dall'ISTAT e completamente Open Source



<https://github.com/SDMXISTATTOOLKIT>

Grazie per l'attenzione

• CARLO BOSELLI | carlo.boselli@istat.it