

Roma, 17 giugno 2022

Forum PA 2022: Talk «La piattaforma Open Source dell'Istat per la diffusione di dati statistici»

LA NUOVA PIATTAFORMA ISTAT A SUPPORTO DEI PROCESSI DI DIFFUSIONE

Sommario

- La modernizzazione delle architetture di diffusione web ISTAT
- Perché lo sviluppo di un toolkit open source
- Perché SDMX
- Principi di progettazione e sviluppo della nuova piattaforma di diffusione
- L'SDMX Istat Toolkit
- Conclusioni

La modernizzazione delle architetture di diffusione web ISTAT

- **Obiettivi:** standardizzazione ed industrializzazione dei processi e dei servizi di diffusione web
- **Strategia:** massimizzazione del riuso di software (tool, librerie, sistemi e servizi web) sviluppato da altre organizzazioni nonché della metainformazione statistica
- **Risultati attesi:**
 - Riduzione dei costi e degli effort di sviluppo nel medio-lungo periodo
 - Soluzioni software durevoli nel tempo
 - Interoperabilità semantica
 - Offrire agli enti del Sistema Statistico Nazionale (SISTAN) una piattaforma «certificata» per favorire una diffusione di qualità dei dati statistici



Perché lo sviluppo di un toolkit open source

- Sono state analizzate in ISTAT *differenti opzioni per l'implementazione della nuova piattaforma di diffusione*: soluzioni commerciali, tool disponibili nell'ambito della comunità statistica, evoluzione di un prototipo sviluppato *in-house*
- La soluzione scelta è stata quella di: **evolvere il prototipo sviluppato in house ed integrarlo con tool sviluppati in altri contesti**
- I criteri utilizzati per lo sviluppo sono stati i seguenti:
 - Progettazione di una **adeguata architettura**
 - Organizzazione dei moduli necessari come un **Toolkit riusabile (gratuito ed open source)**
 - Pianificazione della migrazione dei sistemi corporate** di tipo legacy alla nuova piattaforma di diffusione

SDMX (Statistical Data and Metadata Exchange) è uno standard per la condivisione e l'interscambio di dati statistici sviluppato e mantenuto da 7 organizzazioni internazionali:

- Possiede un **information model** per la strutturazione dei dati e metadati
 - Supporta tutti i **più diffusi formati open data** (XML,JSON,CSV,RDF Data Cube)
- Guida l'armonizzazione dei dati, favorendo l'**interoperabilità semantica**
- **Disponibilità di un numero congruo di tool e librerie open source** sviluppati in particolare dalle grandi organizzazioni internazionali (es. SDMX Reference Infrastructure e librerie SDMX API di Eurostat)

Principi di progettazione e sviluppo della nuova piattaforma di diffusione ISTAT

➤ Approccio “Business-driven”

- Fase pilota con implementazioni prototipali dei tool
- Considerati centinaia di feedback raccolti dagli utenti relativamente alle versioni pilota dei tool

➤ Facilità di installazione e configurazione (modalità “plug&play” e “ready-to-use”)

- Non più di 30 minuti per installazione e configurazione di base
- Skill tecnici facili da reperire in un'organizzazione statistica

➤ Riutilizzo (ove possibile) di moduli e librerie, che incorporano anni di esperienza in SDMX (es. librerie e servizi della SDMX-Reference Infrastructure di Eurostat);

➤ Facilità di uso per le varie tipologie di utenza

➤ Performance adeguate

- Visualizzazione e download di tavole con più di 100.000 celle in pochi secondi
- Benchmark di prestazioni con altre soluzioni sia commerciali che disponibili nell'ambito della comunità statistica

L'SDMX Istat Toolkit ed i processi statistici

- **L'SDMX Istat Toolkit** è un insieme di tool gratuiti ed open source sviluppati da ISTAT che permette la standardizzazione e l'industrializzazione dei processi di diffusione ed interscambio di dati statistici.

Trasformazione

Excel2CSV

Raccolta

- Excel
- CSV (Custom and SDMX)
- SDMX-ML

Diffusione

- GUI (**Data Browser**)
- M2M (**APIs**)
 - SDMX
 - CKAN (DCAT)

Data Manager

- Creazione di cubi di dati multidimensionali
- Alimentazione e gestione dei cubi di dati
- Pubblicazione e gestione delle viste sui dati

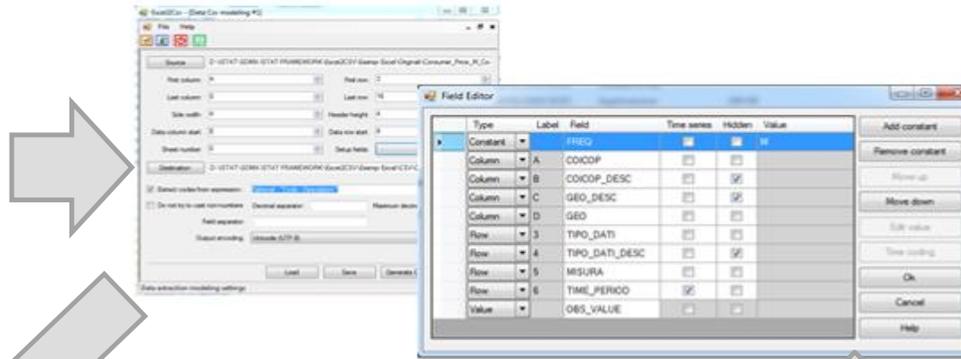
Metadata Manager

- Editing e gestione dei metadata di struttura
- Gestione dei metadata descrittivi
- Registry SDMX
- Glossari
- Cataloghi DCAT

Excel2Csv

- Applicazione desktop che permette di **trasformare una tabella statistica multidimensionale complessa in excel in un file csv “verticale”** (una osservazione per riga). Permette anche di creare file CSV contenenti code list.

Index Type		Consumer price index for the whole nation (Base 2015=100) - monthly					
Measure		[4] Index Number					
Time period		Oct-2011	Nov-2011	Dic-2011	Jan-2012	Feb-2012	Oct-2011
COICOP Rev. Esat	Territory	ITC	ITC	ITC	ITC	ITC	Percep
01 Food and non-alcoholic beverages	Nord-ovest	103,0	104,3	104,3	104,6	105,2	0,6
	Nord-est	103,1	103,9	104	104,3	105,3	0,4
	Centro	102,5	103,4	103,3	103,0	104,6	0,3
02 Alcoholic beverages and tobacco	Nord-ovest	107,2	107,3	107,4	107,5	107,7	3,6
	Nord-est	107	107,1	107,1	107,2	107,3	3,5
	Centro	107,4	107,4	107,4	107,6	107,6	3,7
03 Clothing and footwear	Nord-ovest	102,4	102,7	102,6	102,9	102,7	0,6
	Nord-est	102,4	102,9	102,6	102,9	102,7	0,6
	Centro	102,3	102,3	102,3	102,3	102,4	1,4



Code List (CSV)

```
CODE;NAME
IT;Italy
ITC1;Piemonte
ITC11;Torino
ITC12;Vercelli
ITC13;Biella
ITC15;Novara
ITC16;Cuneo
ITC18;Alessandria
```

Data File (CSV)

```
1 FREQ;COICOP;GEO;TIPO_DATI;MISURA
2 M;01;ITC;9;4;Ott-2011;103,6
3 M;01;ITC;9;4;Nov-2011;104,3
4 M;01;ITC;9;4;Dic-2011;104,3
5 M;01;ITC;9;4;Gen-2012;104,6
6 M;01;ITC;9;4;Feb-2012;105,2
7 M;01;ITC;9;6;Ott-2011;0,6
8 M;01;ITC;9;6;Nov-2011;0,7
9 M;01;ITC;9;6;Dic-2011;0
10 M;01;ITC;9;6;Gen-2012;0,3
11 M;01;ITC;9;6;Feb-2012;0,6
12 M;01;ITC;9;7;Ott-2011;3,3
13 M;01;ITC;9;7;Nov-2011;3,7
14 M;01;ITC;9;7;Dic-2011;3,5
15 M;01;ITC;9;7;Gen-2012;3
16 M;01;ITC;9;7;Feb-2012;3,1
```

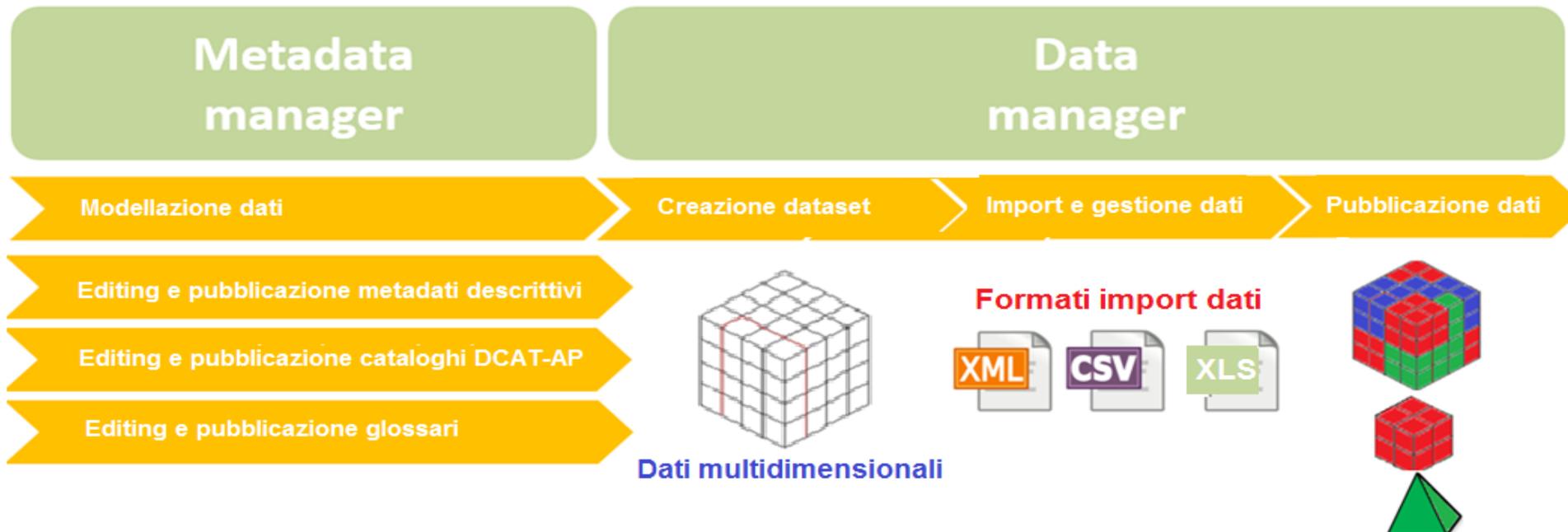
SDMX Data File

```
<?xml version="1.0" encoding="utf-8"?>
<message:StructureSpecificData xmlns:ss="http://www.sdmx.org/SDMX-XML-2009-09-19" xmlns:message="http://www.sdmx.org/SDMX-XML-2009-09-19">
  <message:Header>
    <message:DataSet action="Information" ss:dataScope="Data" ss:structure="ITC" ss:version="1.0" />
  </message:Header>
  <Series FREQ="M" REF_AREA="IT" INDICATOR="CPI_FOI" />
    <Obs TIME_PERIOD="2021-04" OBS_VALUE="110.4"/>
    <Obs TIME_PERIOD="2021-05" OBS_VALUE="110.4"/>
    <Obs TIME_PERIOD="2021-06" OBS_VALUE="110.2"/>
    <Obs TIME_PERIOD="2021-07" OBS_VALUE="110.4" />
  </Series>
  <Series FREQ="M" REF_AREA="IT" INDICATOR="CPI_FOI" />
    <Obs TIME_PERIOD="2021-04" OBS_VALUE="102.1"/>
    <Obs TIME_PERIOD="2021-05" OBS_VALUE="102.3"/>
  </Series>
</message:StructureSpecificData>
```

Meta and Data Manager

➤ Tool web SDMX Based composto di due moduli principali:

- ❑ **Meta(data) Manager:** per la modellazione dei dati e per l'editing e la pubblicazione di: metadati descrittivi, cataloghi DCAT-AP_IT di dati aperti e glossari
- ❑ **Data Manager:** modulo per la creazione di dataset multidimensionali basati sulla modellazione dati SDMX, l'import e la gestione dei dati e la pubblicazione di viste sui dati memorizzati nei dataset



Può connettersi a più ambienti di lavoro e navigare repository di metadati SDMX pubblici di altre organizzazioni

Data Browser

- Applicazione web per la **navigazione, la visualizzazione e lo scarico dei dati da una o più sorgenti**
- Elevate prestazioni grazie ad un avanzato **sistema di caching** gestibile anche da Meta&Data manager
- Possiede una **agevole interfaccia di amministrazione e configurazione del sistema**

A
M
M
I
N
I
S
T
R
A
Z
I
O
N
E

SWITCH DEI NODI

FUNZIONI DI RICERCA

NAVIGAZIONE DATI

VISUALIZZAZIONE DATI

DASHBOARD

CONDIVISIONE

PORTAFOGLIO UTENTI

Tavole

Grafici

Mappe

Criteri

Pivoting

Download

Download

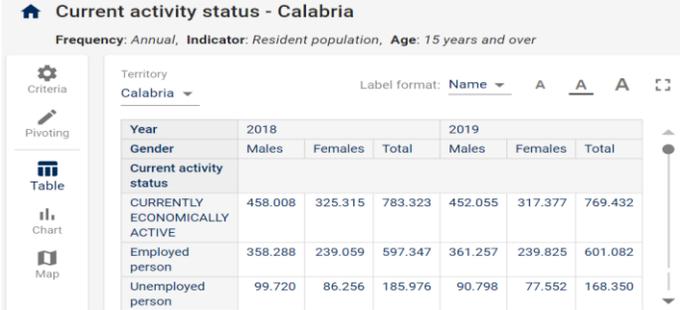
XML

EXCEL

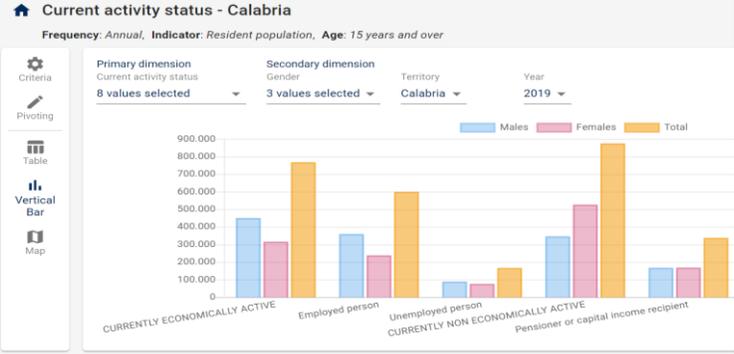
CSV

SDMX

JSON



Year	2018			2019		
Gender	Males	Females	Total	Males	Females	Total
Current activity status						
CURRENTLY ECONOMICALLY ACTIVE	458.008	325.315	783.323	452.055	317.377	769.432
Employed person	358.288	239.059	597.347	361.257	239.825	601.082
Unemployed person	99.720	86.256	185.976	90.798	77.552	168.350




Conclusioni

- La modernizzazione di una architettura di diffusione non può prescindere da una precisa **strategia e visione corporate**
- Tale strategia, per essere efficace, deve essere **sostenibile, tenere conto delle esigenze dell'utenza ed essere in linea con standard, best practices e linee guida a livello nazionale ed internazionale**
- La **sostenibilità** deve essere valutata sui seguenti **parametri**: *funzionalità disponibili, usabilità, curva di apprendimento, facilità di installazione ed evoluzione delle piattaforme, skill necessari*
- Sulla base degli aspetti di sostenibilità, **deve essere effettuata una opportuna pianificazione delle milestones relative agli sviluppi del software ed alla messa in produzione** dei nuovi sistemi di diffusione
- **Concetti chiave** nella progettazione e sviluppo software: *riuso (tool, librerie), service orientation e modularità* delle implementazioni software

Link utili

- **SDMX Istat Toolkit:** <https://sdmxistattoolkit.github.io>
- **SDMX.org:** <http://www.sdmx.org>
- **Sistan Hub:** <https://sistanhub.istat.it>
- **Censimento permanente della Popolazione:** <https://esploradati.censimentopopolazione.istat.it>

Grazie dell'attenzione

ALESSIO CARDACINO | alcardac@istat.it