



INTERSTAT

Open Statistical Data
Interoperability Framework

Il progetto **INTERSTAT**: obiettivi e risultati

ISTAT Webinar - 25 gennaio 2024

Martino Maggio
Engineering Ingegneria Informatica SpA



Co-financed by the Connecting Europe
Facility of the European Union

The contents of this publication are the sole responsibility of INTERSTAT consortium
and do not necessarily reflect the opinion of the European Union



www.cef-interstat.eu

Dati statistici e interoperabilità

- I dati statistici assumono un ruolo sempre più importante nella definizione e implementazione delle policy nell'ambito sociale economico e amministrativo.
- Per la loro rilevanza, i dati statistici fanno parte degli «High Value Dataset» (HVDs), le categorie di dati di maggior importanza, definite all'interno della Open Data Directive e del suo regolamento di attuazione
- E' fondamentale l'aspetto dell'interoperabilità dei dati statistici tra i diversi Paesi europei in maniera da poter mettere a disposizione delle comunità scientifiche, cittadini e amministrazioni, dati aperti, riusabili per la creazione di nuovi servizi cross-border.
- Si aggiunge a questi aspetti la necessità, più generale, di superare il problema della frammentazione degli open data: evitare che ogni ente utilizzi piattaforme e standard differenti per la loro pubblicazione ostacolando la ricerca e fruizione dei dati.



Progetto INTERSTAT

CEF Telecom call 2019 - Public Open Data

Inizio: Settembre 2020

Fine: Agosto 2023

Durata: 36 mesi



CEF Digital
Connecting Europe

«L'obiettivo di INTERSTAT è quello di sviluppare soluzioni per l'interoperabilità, utilizzando risorse tecniche e ontologie comuni, tra i portali statistici nazionali e il portale europeo dei dati e l'implementazione di servizi cross-border che riutilizzino i dataset statistici europei aperti.»

Partner del progetto



Co-financed by the Connecting Europe
Facility of the European Union

The contents of this publication are the sole responsibility of INTERSTAT consortium
and do not necessarily reflect the opinion of the European Union

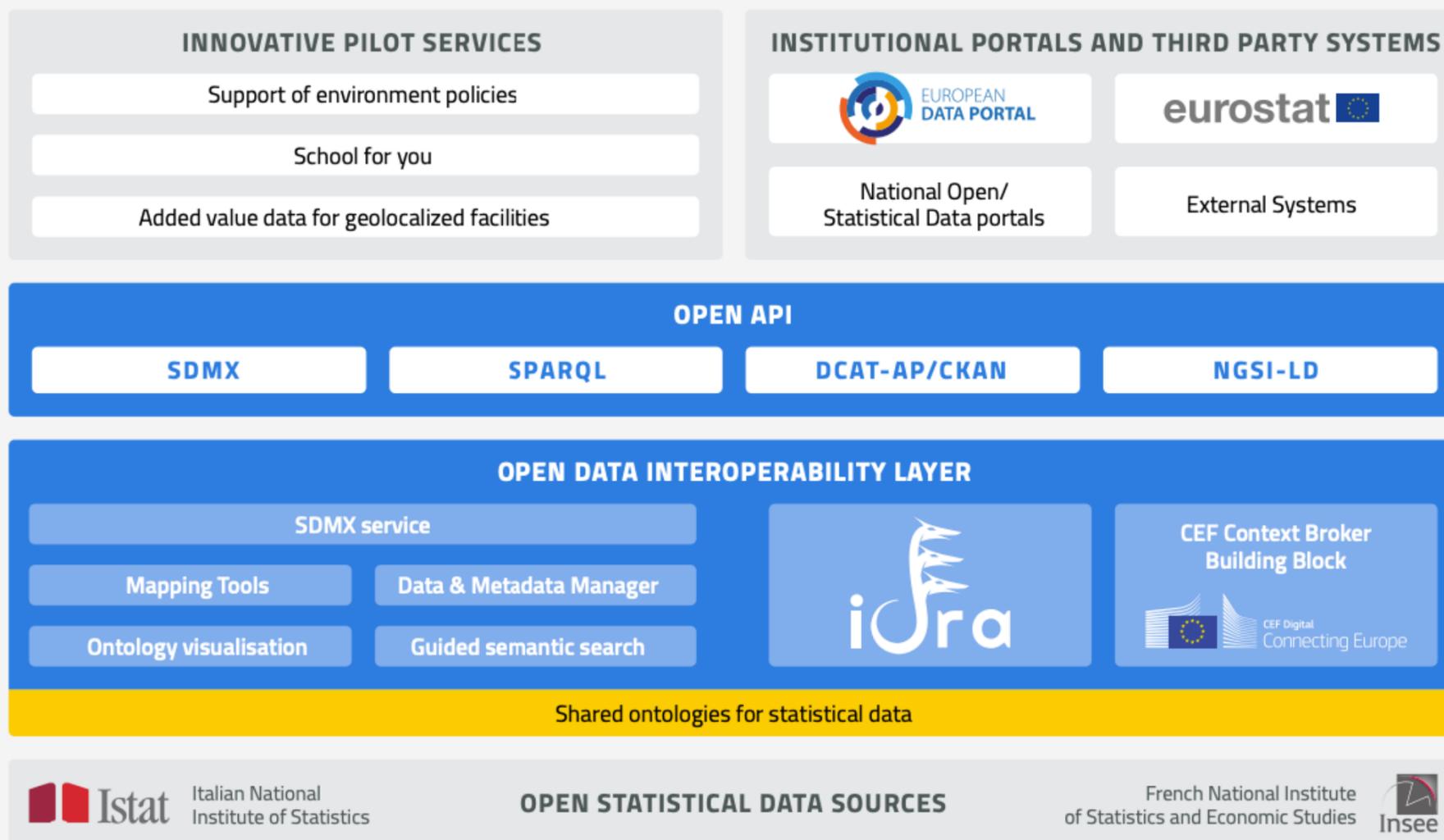


Obiettivi

1. Permettere **l'interoperabilità** tra i diversi portali statistici nazionali e il portale europeo dei dati attraverso l'adozione di standard e strumenti per automatizzare l'armonizzazione dei metadati e la pubblicazione dei dati
2. Fornire standard, metodologie e strumenti per raggiungere **l'armonizzazione** dei dati nel campo dei (Linked) Open data statistici tra i diversi istituti statistici nazionali
3. Fornire interfacce tecniche (API) per un **riutilizzo standard** e semplice delle informazioni statistiche
4. Fornire strumenti per **semplificare la visualizzazione e l'analisi** dei dati statistici per gli utenti finali (non tecnici)
5. Validare le soluzioni tecniche, sviluppate nel progetto, attraverso l'implementazione e la sperimentazione di servizi cross-border per gli utenti finali riutilizzando **dati statistici armonizzati** in combinazione con open data provenienti dal Portale europeo dei dati e altri portali nazionali di dati aperti.



Architettura del framework



Il framework **open source** di INTERSTAT fornisce:

- Funzionalità per **l'armonizzazione dei Linked Open Statistical Data (LOSD)**, (forniti da Istat e Insee), attraverso l'adozione di modelli dati comuni e la fornitura di strumenti specifici per la mappatura, l'interrogazione e la visualizzazione dei dati, in conformità allo standard SDMX.
- La piattaforma **Idra**, per la federazione e l'armonizzazione di dataset aperti provenienti da fonti eterogenee e la loro fornitura attraverso interfacce standard e modelli di metadati (es. DCAT-AP)
- CEF **Context Broker** Building Block che permette l'accesso ai LOSD attraverso i modelli e le API NGSI-LD
- Una serie di API aperte basate su diversi standard che permettono l'accesso e la condivisione dei LODS attraverso diversi portali istituzionali di Open Data in Europa (in particolare il portale europeo dei dati), ma anche sistemi di terze parti.



Un framework open e accessibile

The screenshot shows the main page of the Interstat Framework. At the top, there is a navigation bar with links for Home, About, Tools, Information, Help, and Partners. Below the navigation bar is a large blue banner with the text "Open Statistical Data Interoperability Framework". The main content area is a grid of tool cards. Each card features a logo, the tool name, a brief description, and a small globe icon. The tools shown include iOra (Open Data Federation Platform), FIWARE (CEF Orion Context Broker), GraphDB (Tool for Data Publication), Adminer (Database management in a single PHP file), and Derilinx (Cube Visualizer and Olap Browser).

interstat.eng.it

The screenshot shows the "Information about Tools" page. It features a search bar at the top right. Below the search bar is a table with the following columns: Name, Documentation, Main feature, Version, Main GSBPM, and Owner Partner. The table lists several tools, including Bauhaus, CEF Context Broker, CKAN Portal, Cube Visualizer, and Data Browser. A sidebar on the left contains a "Services" menu with options like GSBPM, GSBPM Processes, Specify Needs, Design, Build, Collect, Process, Analyse, Disseminate, and Evaluate.

| Name | Documentation | Main feature | Version | Main GSBPM | Owner Partner |
|--------------------|---------------|---|----------|-------------------------------------|---------------|
| Bauhaus | | Management of code lists, data structure definition and DCAT-AP catalogue entries | 3.0.7 | 5.4. Edit and impute | Insee |
| CEF Context Broker | | Allows for the management and requesting context of information in a structured manner based on LD standards following the ETSI NGSI-LD specification | Jan 2021 | 7.2. Produce dissemination products | FIWARE |
| CKAN Portal | | CKAN portal for the publication of data produced within the project. Data will be entered within the INTERSTAT organization present in the portal. | 2.9.4 | 7.2. Produce dissemination products | |
| Cube Visualizer | | Data analysis and visualisation tool that creates graphical representations of an RDF data cube's one-dimensional slices | Apr 2019 | 6.3. Interpret and explain outputs | Derilinx |
| Data Browser | | Interacts with SDMX web services allowing data-users to browse, present and visualize datasets | 1.1.1 | 7.2. Produce dissemination | ISTAT |



Alcuni tool del framework INTERSTAT...

European Commission
eurostat Your key to European statistics

Request information

| Property | Value |
|-------------------------------|-------------------------------|
| Root URL | http://interstat.opsi-lab.it/ |
| Requested Host/IP | interstat.opsi-lab.it |
| Port | 80 |
| Is usable from external users | True |

Endpoints

| Service Name | Endpoint path | Namespace | WSDL link | XML Schema path |
|------------------------------------|--|---|-----------|-----------------|
| SDMX v2.0 with Eurostat extensions | http://interstat.opsi-lab.it/NSIStdV20Service | http://ec.europa.eu/eurostat/sri/service/2.0/extended | WSDL | SDMXMessage.xsd |
| Standard SDMX v2.0 | http://interstat.opsi-lab.it/NSIStdV20Service | http://ec.europa.eu/eurostat/sri/service/2.0 | WSDL | SDMXMessage.xsd |
| Standard SDMX v2.1 | http://interstat.opsi-lab.it/SdmxService | http://www.sdmx.org/resources/sdmxml/schemas/v2_1/webservices | WSDL | SDMXMessage.xsd |
| SDMX v2.1 Registry | http://interstat.opsi-lab.it/SdmxRegistryService | http://www.sdmx.org/resources/sdmxml/schemas/v2_1/webservices | WSDL | SDMXMessage.xsd |

RESTful API

| Service name | Resource base | WADL link |
|--------------|------------------------------------|--|
| RESTful API | http://interstat.opsi-lab.it/rest/ | http://interstat.opsi-lab.it/rest/application.wadl |

Remarks

This page appears to be have been requested using an external (internet) host name or IP address. External should be able to access the endpoints configured for NSI Web Service using one of the above URLs.

Copyright (c) 2009 by the European Commission, represented by Eurostat.

insight

Measure: Value Free dimension: Age Refresh data

Country: France Gender: F NutsRegion: FR Year: 2011 measureType: Value Chart type

Metadata management application

M&D M

- Concept Schemes
- Codelists
- Data Structure
- Dataflows
- Category Schemes
- Agency Schemes
- Content Constraints

Agencies

Endpoint

Annotations

Proxy

* SDMX WS Endpoint type: SOAP

* SDMX WS Endpoint: https://localhost/NSI_WS/SdmxReg

Remote SDMX WS username:

Remote SDMX WS password:

Concepts - Search

Label...

Advanced search

1131 results

- Accédant (Propriétaire)
- Accidents corporels de la circulation
- Accroissement total de population
- Accueils avec hébergement (mineurs)
- Accueils collectifs de mineurs à caractère éducatif
- Accueils de scoutisme (mineurs)
- Accueils sans hébergement (mineurs)

+ New

Export

Publish

Datalift

Repository: Published data Response format: HTML RDF/XML N3/Turtle N-Triples TRIG TRIX

Query: SELECT DISTINCT ?s WHERE { ?s ?p ?o . }

Predefined queries: Resources Types Graphs All entries

Max. results: 1000 Display literal types

More information about Datalift

RQL React

Summary

Establishments and populations

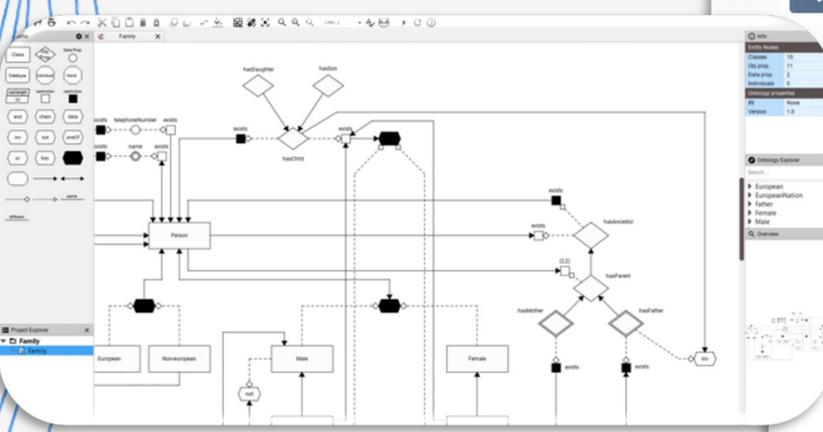
- Population near an establishment
- Population near a type of establishment

Building permits and populations

- "Île de France" départements

Pop 5

- 15 year or more population
- Departmental populations: overview



French_HC55_ds

Table dimensions: Row: Age Column: Country

Filter: Measure: Value Gender: F Year: 2011 measureType: Value NutsRegion: FR

Show table

Rotate table: Age ↔ Country

| TOTAL | Y15-29 | Y30-49 | Y50-64 | Y65-84 | Y_GE85 | T15 |
|-------|--------|--------|--------|--------|--------|-----|
| | | | | | | |

Istat StatKit Data Browser

Torna alla homepage del Nodo

Categorie

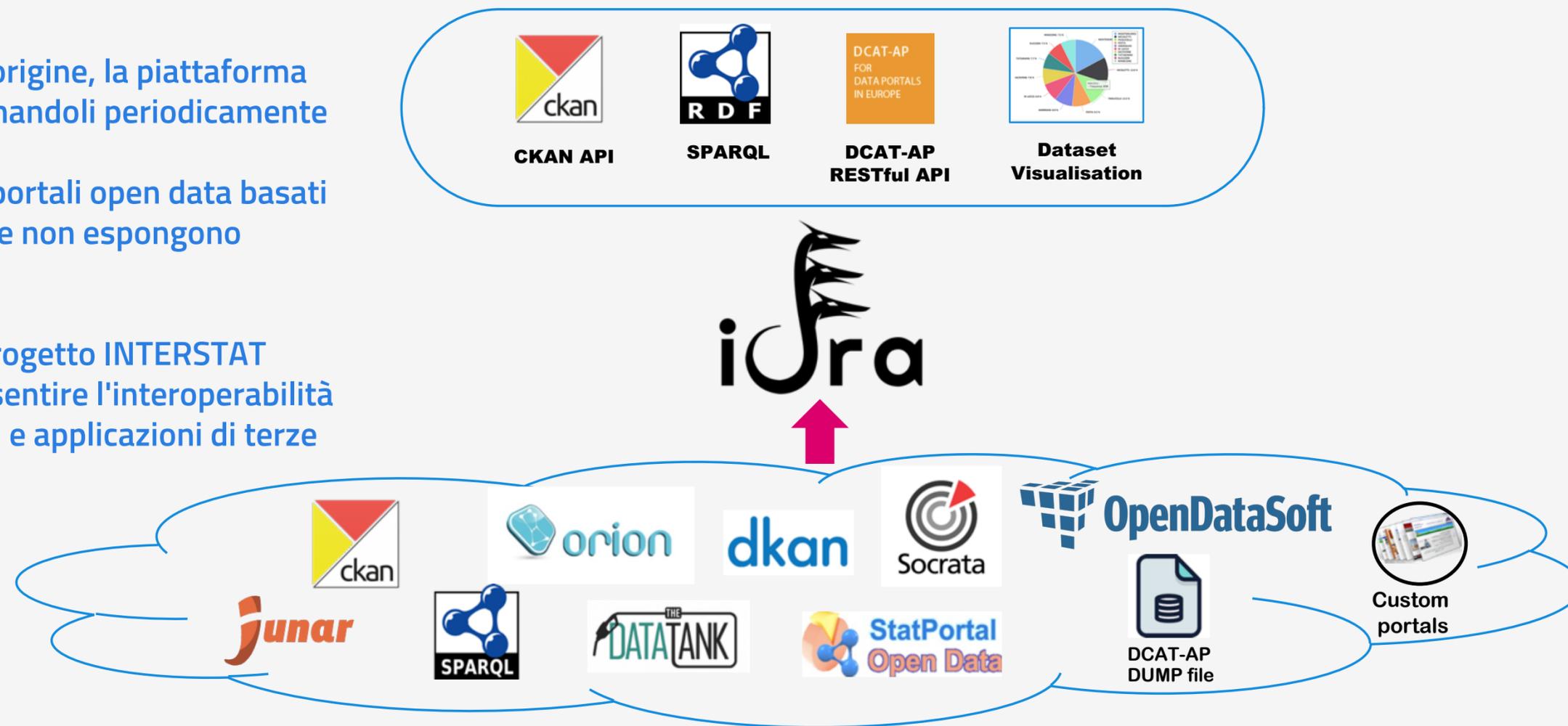
- Imprese
- Prezzi
- Commercio con l'estero e internalizzazione
- Pubbliche amministrazioni e istituzioni private
- Servizi
- Industria e costruzioni
- Agricoltura
- Conti nazionali
- Lavoro e retribuzioni
- Vita quotidiana e opinione dei cittadini
- Cultura, comunicazione, viaggi
- Istruzione e formazione
- Assistenza e previdenza
- Salute e sanità
- Condizioni economiche delle famiglie e disuguaglianze

I.Stat

ESPLORA

Alcuni strumenti per l'interoperabilità: Idra - Open Data Federation Platform

- Idra è una piattaforma open source che è in grado di fornire un unico punto di accesso agli open data di pubbliche amministrazioni o enti privati provenienti da sorgenti e portali basati su tecnologie eterogenee
- Gli open data rimangono nei portali di origine, la piattaforma importa e armonizza i metadati aggiornandoli periodicamente
- La piattaforma è capace di accedere a portali open data basati su differenti tecnologie anche quelli che non espongono specifiche API
- Idra è uno degli strumenti chiave del progetto INTERSTAT insieme al CEF Context Broker per consentire l'interoperabilità tra tecnologie eterogenee di dati aperti e applicazioni di terze parti



GitHub: <https://github.com/OPSILab/Idra>

Developers Italia: <https://developers.italia.it/it/software/opsilab-idra-015316>

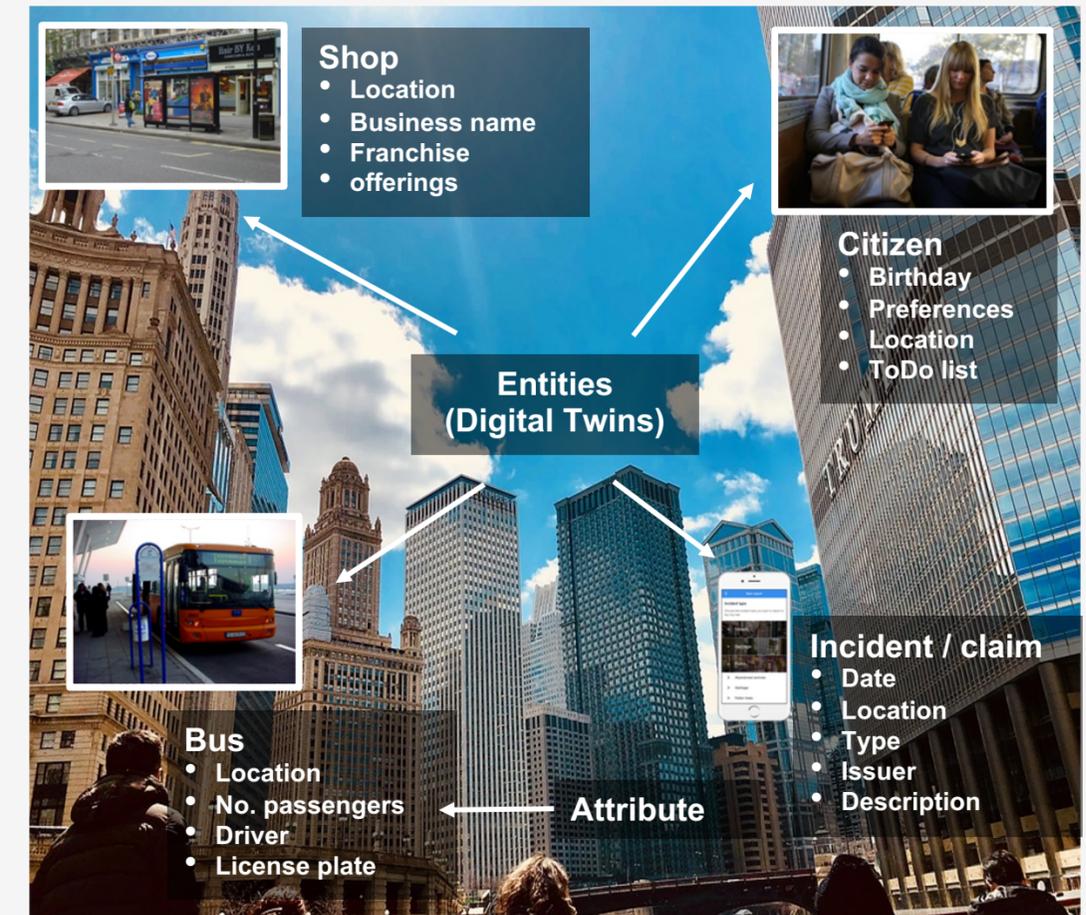


Co-financed by the Connecting Europe
Facility of the European Union

The contents of this publication are the sole responsibility of INTERSTAT consortium
and do not necessarily reflect the opinion of the European Union



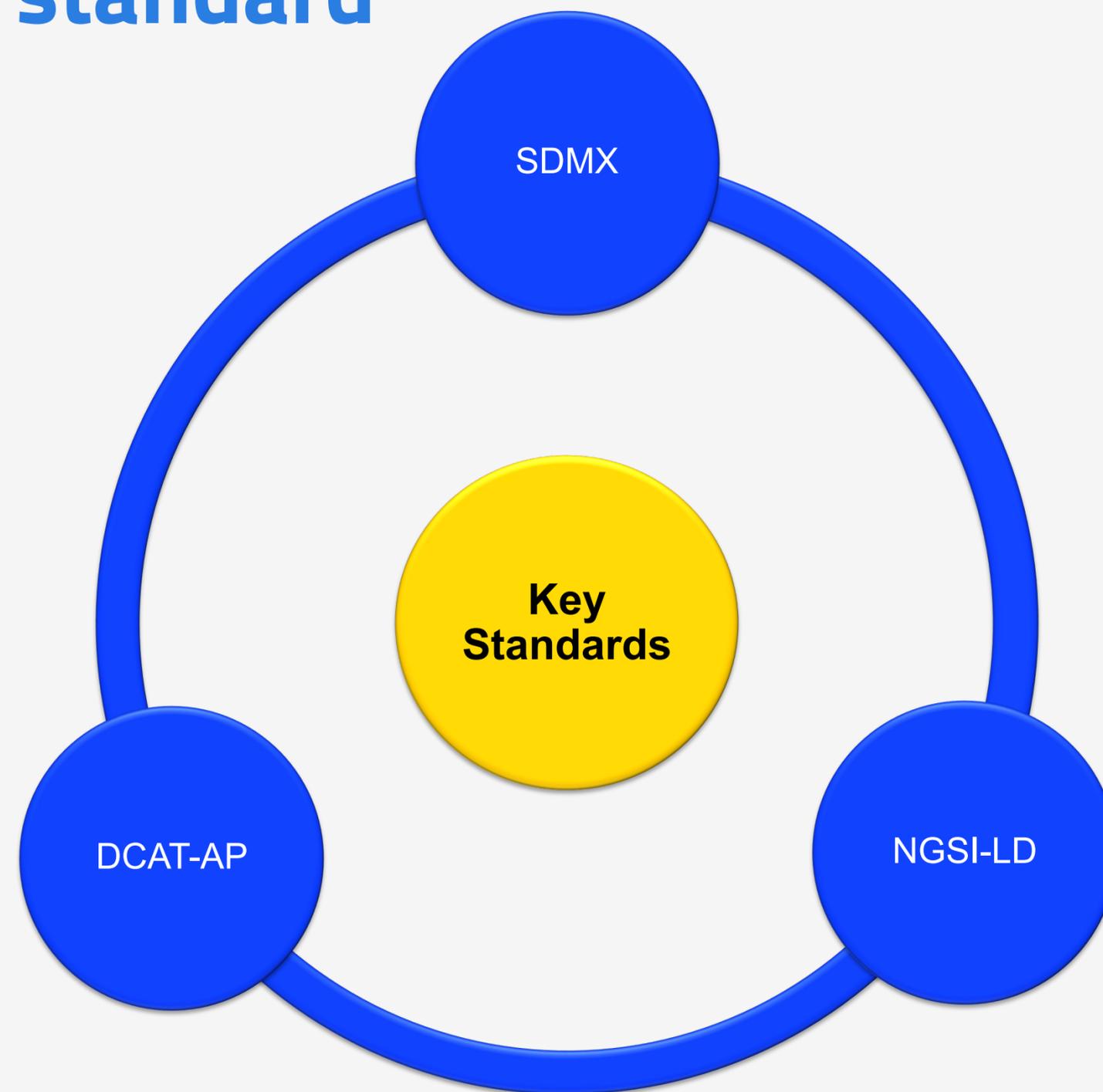
Alcuni strumenti per l'interoperabilità: CEF Context Broker



<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Context+Broker>

Interoperabilità attraverso gli standard

- **SDMX: Statistical Data and Metadata eXchange** è lo standard internazionale per lo scambio dei dati e metadati tra gli istituti di statistica.
- **DCAT-AP: il profilo applicativo basato sul vocabolario DCAT, una specifica del W3C per descrivere i dataset del settore pubblico in Europa.**
- **ETSI NGSI-LD: standard europeo per la gestione delle informazioni di «contesto» ed in tempo reale**



Parser SDMX/NGSI-LD

È stato sviluppato un parser compatibile con DCAT-AP v.2.0.1 e statDCAT-AP v1.0.1 per tradurre i dati SDMX rappresentati nel formato file Turtle RDF in formato JSON-LD, in modo da essere conformi alle specifiche ETSI NGSI-LD API.

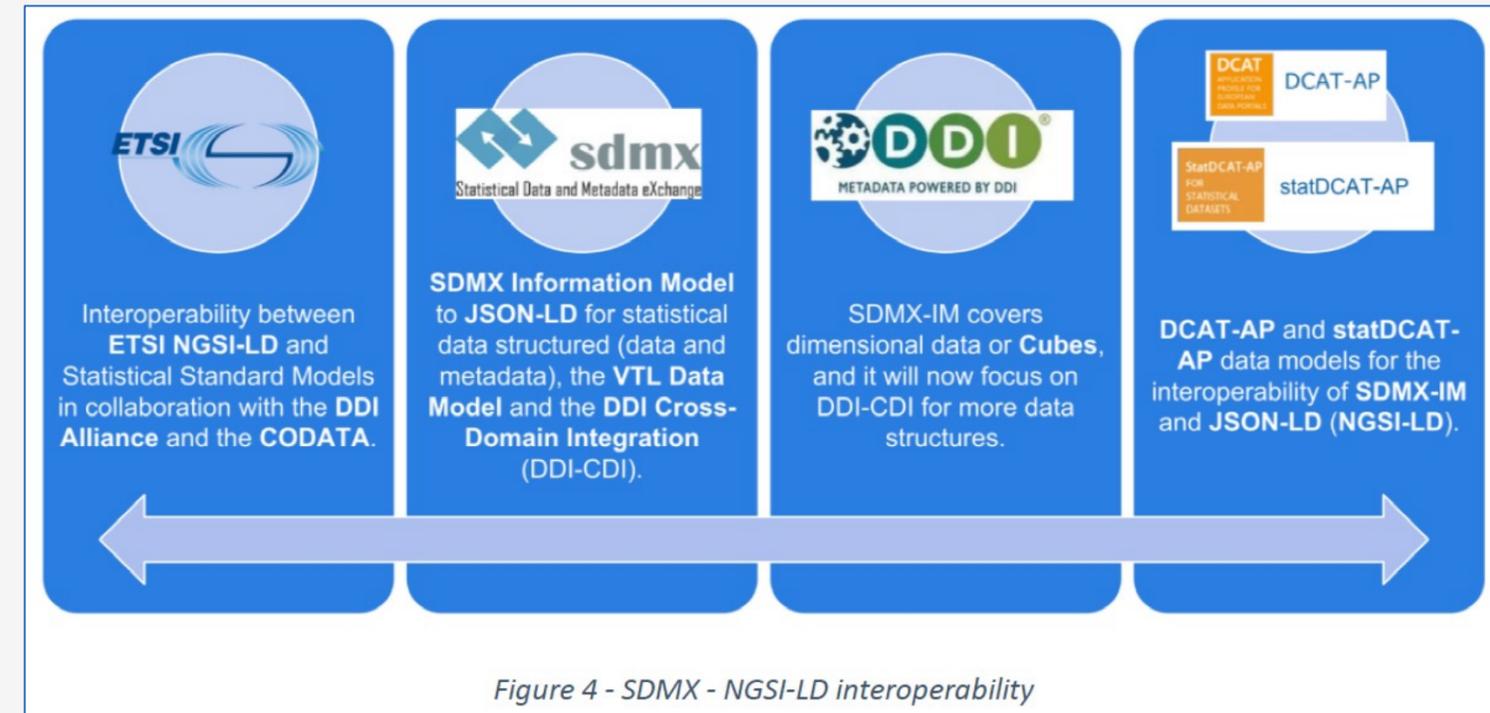
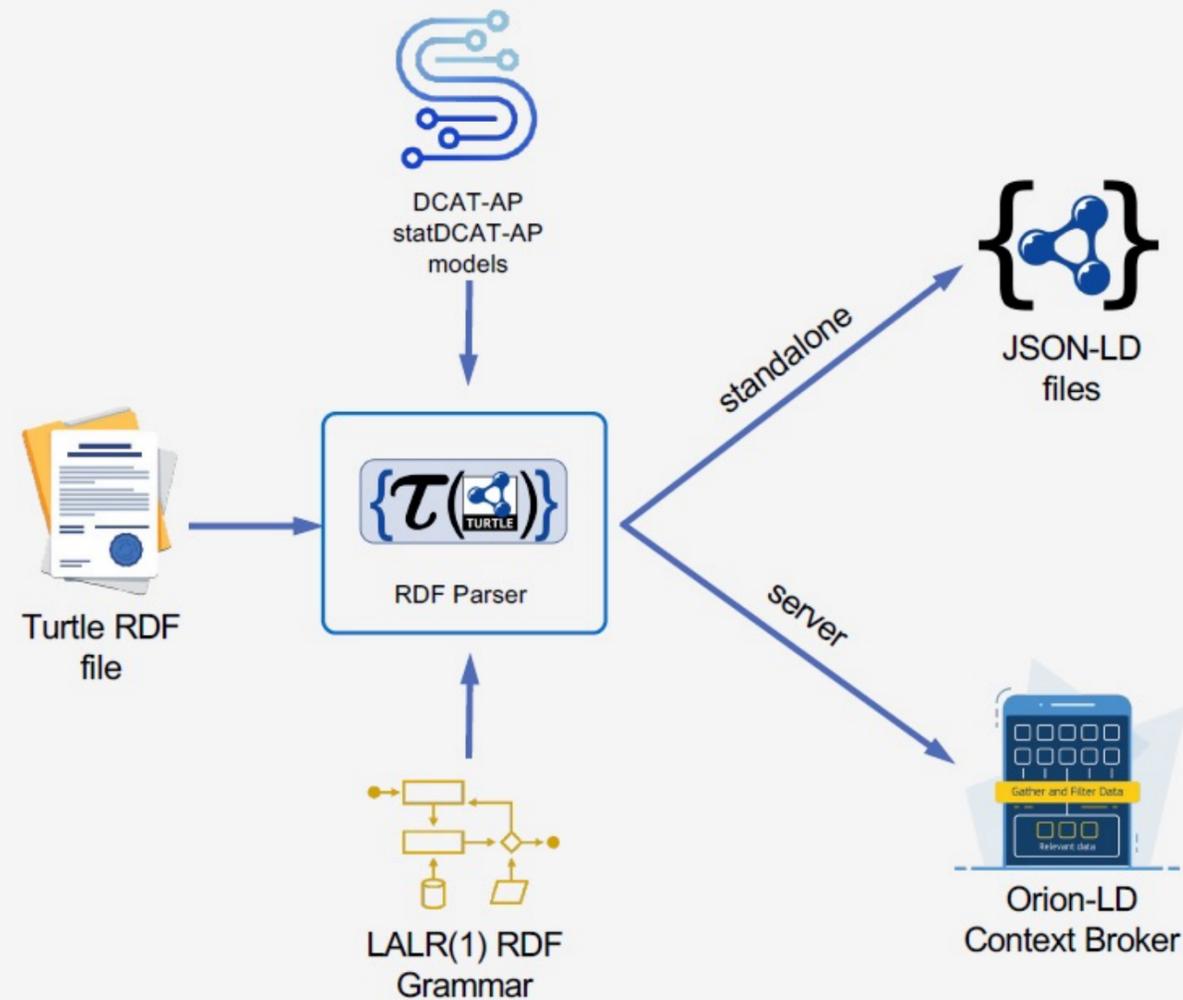


Figure 4 - SDMX - NGSI-LD interoperability

Interoperabilità tra ETSI NGSI-LD (JSON-LD) e i modelli statistici standard (SDMX-IM) attraverso il modello di dati statDCAT-AP

<https://github.com/INTERSTAT/loTAgent-Turtle>

<https://github.com/smart-data-models/dataModel.STAT-DCAT-AP>

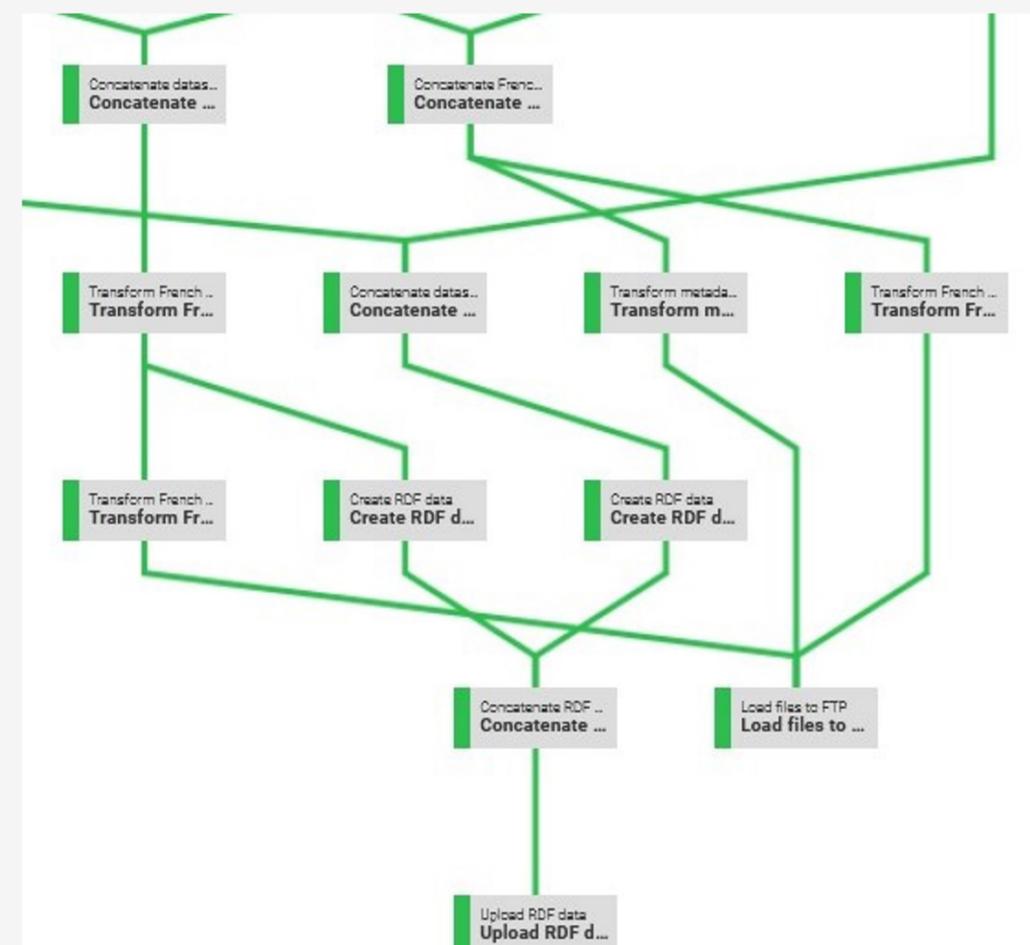
Data Pipelines: approccio ETL

Le pipeline di dati implementate per le applicazioni pilota si basano su due approcci diversi, che convergono alla fine del flusso di lavoro per la produzione dei dati di output:

Approccio ETL: che genera triple RDF da un dataset CSV attraverso pipeline basate su Python/Prefect o VTL.

Principali vantaggi:

- Open source: il codice, sviluppato con strumenti aperti, è disponibile nel repository GitHub di INTERSTAT.
- Massima automazione, per evitare attività manuali, risparmiare tempo e migliorare la tracciabilità.
- Riproducibilità, grazie all'automazione e alla documentazione del codice.
- Efficienza, aumentata dall'esecuzione della pipeline in un ambiente distribuito.

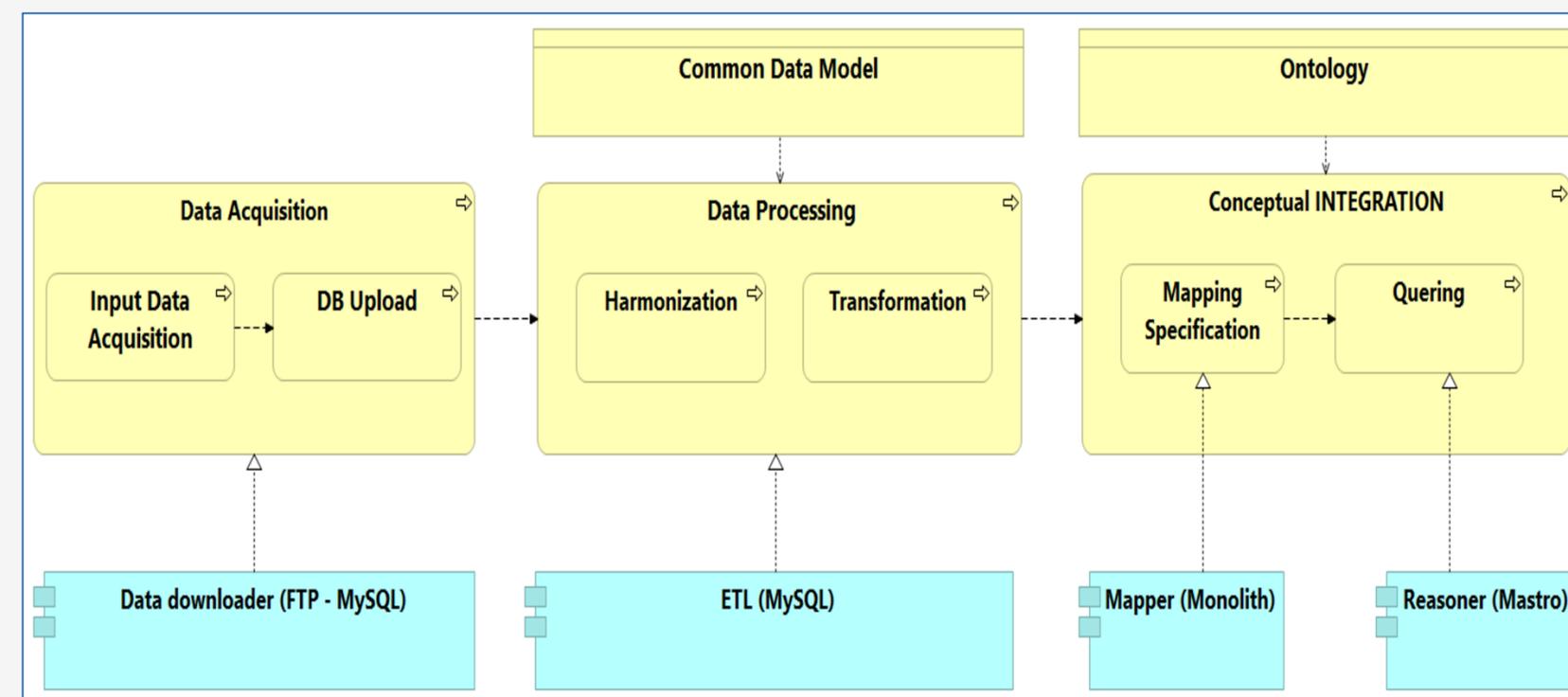


Data pipelines: approccio Domain Knowledge

Approccio Domain Knowledge, basato sulla descrizione del dominio di interesse attraverso un'ontologia e sulla definizione di un Common Data Model per collegare fonti di dati eterogenee con i concetti dell'ontologia.

I principali vantaggi della Domain Knowledge pipeline:

- Definizione formale e chiara dei concetti target e dei relativi metadati
- Automatic reasoning
- Interoperabilità cross domain by design
- Disaccoppiamento tra la struttura dei dati e la loro semantica
- Approccio incrementale e gestione semplificati dei dati e dei metadati
- Collegamento più semplice di nuove fonti di dati esterne



INTERSTAT: applicazioni pilota

S4Y – The school for you

Permettere alle famiglie, che devono scegliere la scuola più adatta ai propri figli, di accedere ad informazioni integrate relative ai dati provenienti dal portale web delle diverse scuole di una zona di un comune scelto e ai dati provenienti dal Censimento della popolazione e delle abitazioni relativi alla stessa area geografica.

Geolocalized Facilities

Fornire informazioni riguardanti eventi e infrastrutture geolocalizzate per fornire un supporto decisionale all'utente: potrebbe essere utilizzata da un utente che visita un luogo che non conosce, e si chiede dove si trovano le strutture più vicine di diverso tipo o quali eventi sono programmati nei vicini stadi, teatri o luoghi culturali. Ma può anche essere usata da una persona che deve effettuare un possibile investimento a livello locale.

Support for Environment Policies

Supportare i policy maker locali che devono prendere decisioni sulle politiche ambientali da applicare in una città. In particolare, i policy maker locali possono beneficiare di dataset integrati derivanti da: (i) dati dei sensori relativi all'inquinamento atmosferico e (ii) dati statistici relativi alla caratterizzazione demografica delle aree della città.

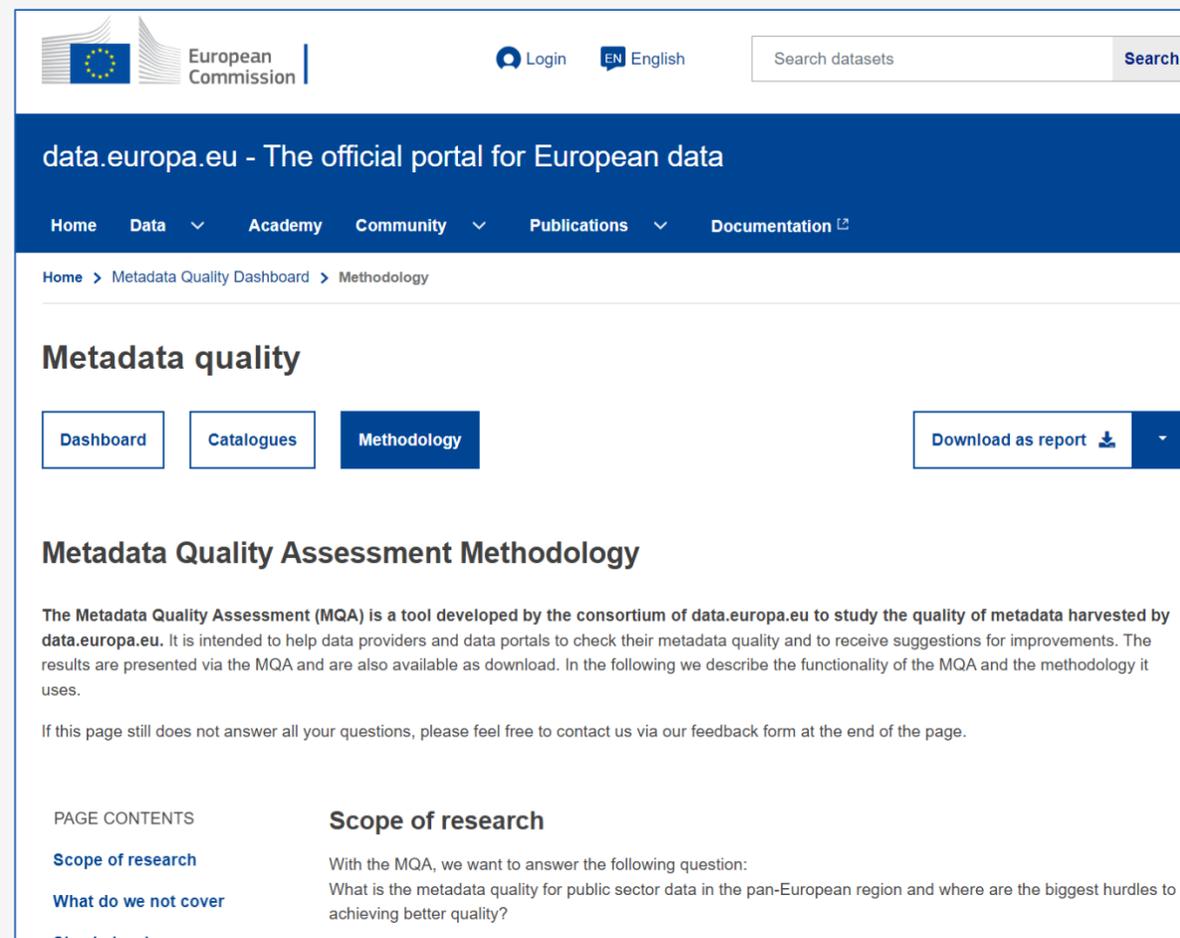
The image displays three screenshots of the INTERSTAT pilot applications. The top screenshot shows the 'S4Y' interface for Italy (IT) with a map of Rome and a facility code RMMM8FR013. The middle screenshot shows the 'Geolocalized Facilities' interface for France (FR) with a map of Cannes and a list of facility types. The bottom screenshot shows the 'Support for Environment Policies' interface for France (FR) and Italy (Alessandria) with a map of the Mediterranean region and population data for the Y25-29 age group.

| Geographic areas | Total population in the selected French territory for the Y25-29 Age Group | Total population in the selected Italian territory for the Y25-29 Age Group |
|--------------------------------------|--|---|
| Paris (France) / Alessandria (Italy) | 912232 | 18876 |



Compliance con Metadata Quality Assurance (MQA)

- I dataset di INTERSTAT più rilevanti sono stati pubblicati sul Portale Europeo dei Dati (data.europa.eu) e validati in base alla metodologia Metadata Quality Assurance secondo i principi FAIR. (**Findability, Accessibility, Accessibility, Reusability**)



European Commission

data.europa.eu - The official portal for European data

Home > Metadata Quality Dashboard > Methodology

Metadata quality

Dashboard Catalogues Methodology Download as report

Metadata Quality Assessment Methodology

The Metadata Quality Assessment (MQA) is a tool developed by the consortium of data.europa.eu to study the quality of metadata harvested by data.europa.eu. It is intended to help data providers and data portals to check their metadata quality and to receive suggestions for improvements. The results are presented via the MQA and are also available as download. In the following we describe the functionality of the MQA and the methodology it uses.

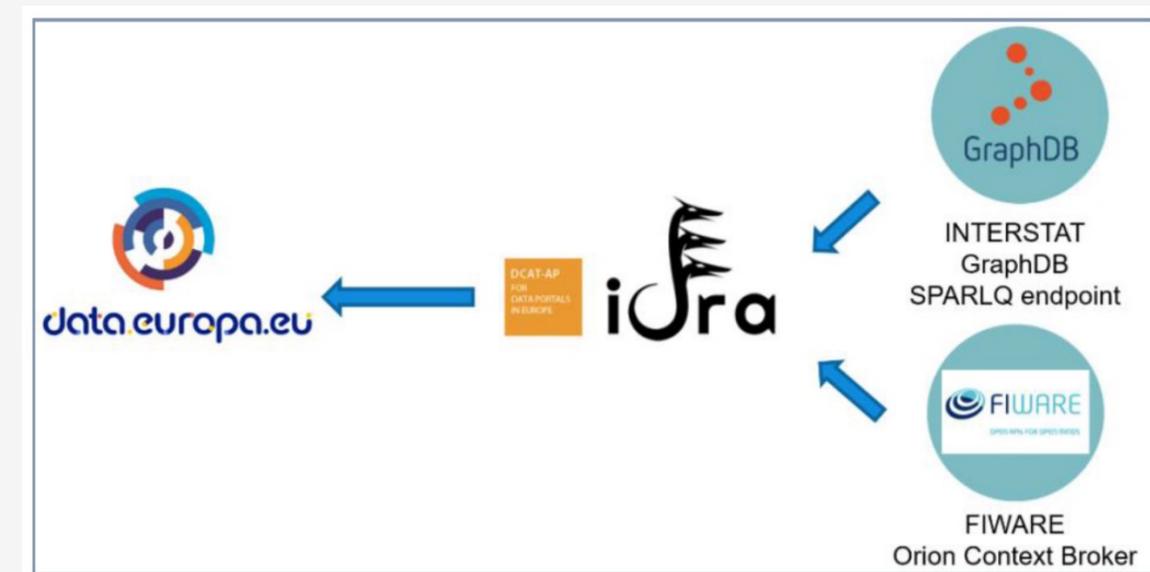
If this page still does not answer all your questions, please feel free to contact us via our feedback form at the end of the page.

PAGE CONTENTS

- Scope of research
- What do we not cover
- Shacl levels

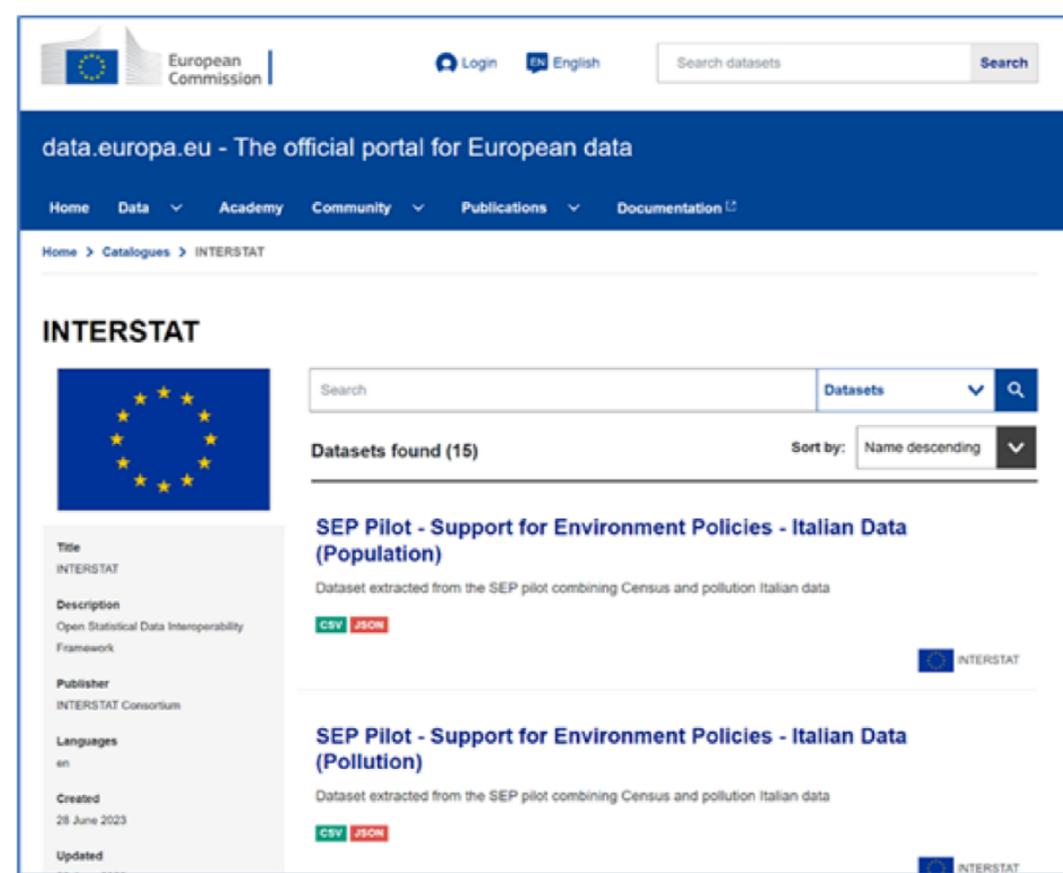
<https://data.europa.eu/mqa/methodology?locale=en>

Il processo di harvesting dei dataset INTERSTAT è stato effettuato utilizzando lo strumento Idra in combinazione con lo SPARQL-endpoint GraphDB e il broker Orion Conxtext che raccolgono rispettivamente i dataset RDF e NGS-LD prodotti durante il progetto.

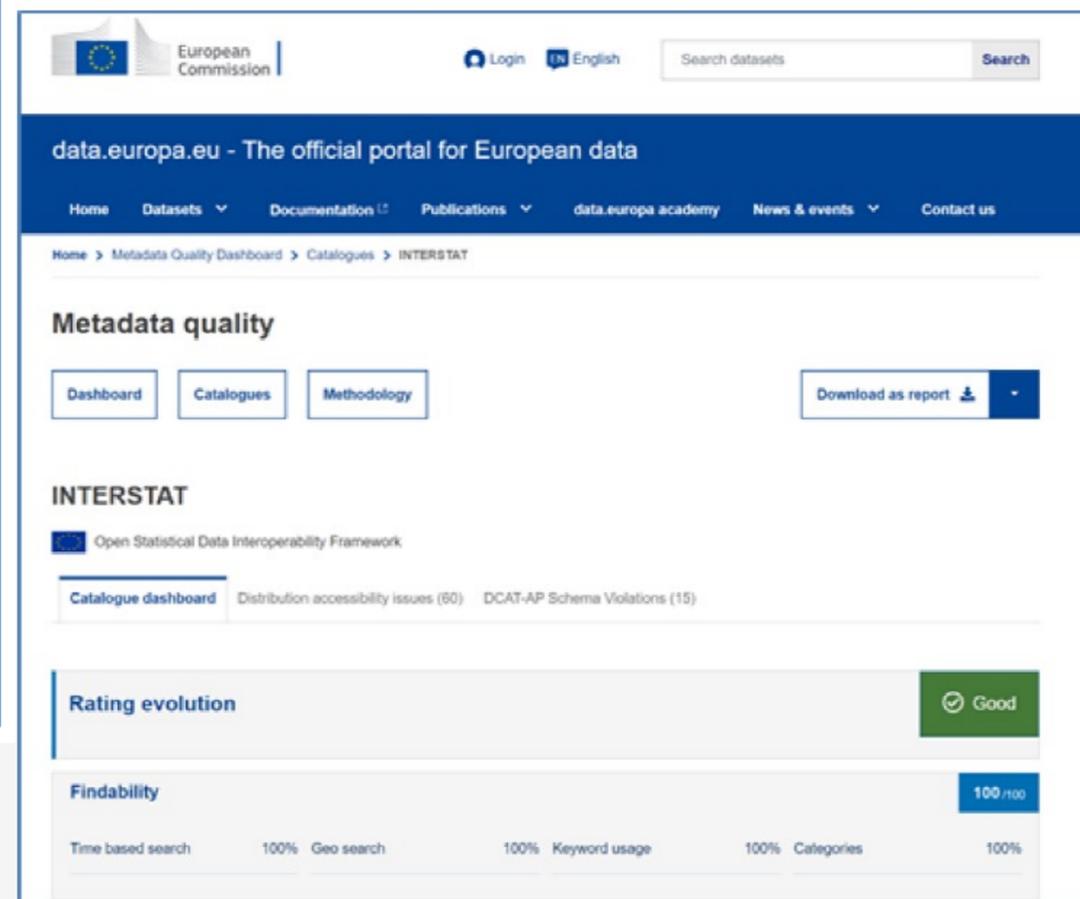


Compliance con Metadata Quality Assurance (MQA)

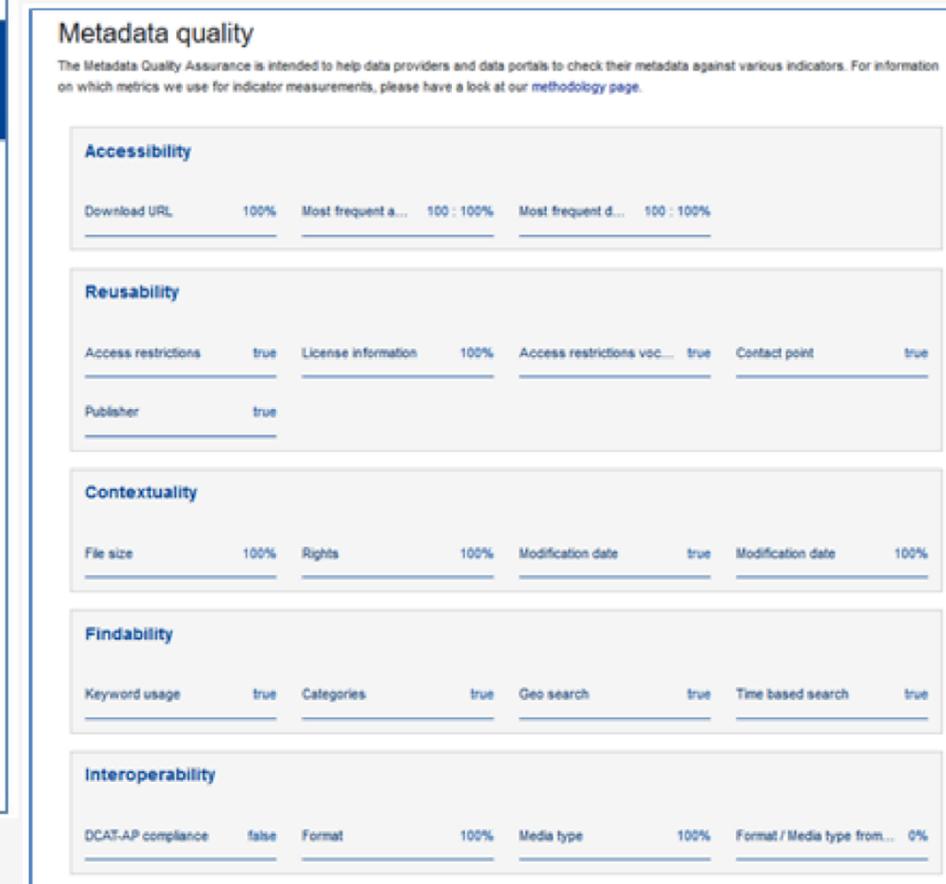
Catalogo INTERSTAT pubblicato su data.europa.eu¹



Valutazione per l'intero Catalogo INTERSTAT



Valutazione per ogni dataset del Catalogo INTERSTAT



¹ <https://data.europa.eu/mqa/catalogues/interstat/?locale=en>



Principali risultati ottenuti

- Analisi e report sulle barriere legali e tecniche in Europa che possono impedire il riutilizzo e lo sfruttamento dei (Linked) Open data statistici.
- Framework di tool riusabili orientati alla produzione, interoperabilità e riutilizzo di dataset aperti nell'ambito statistico (ma applicabili anche ad altri domini)
- Definizione di differenti approcci basati su data pipelines semi-automatizzate per la produzione dei dataset
- Validazione degli open dataset prodotti secondo la metodologia MQA e pubblicazione sul Portale Europeo dei Dati
- Attività di ricerca e sviluppo di prototipi software per l'interoperabilità tra standard nell'ambito statistico e dei dati aperti di contesto
- Validazione dei risultati attraverso lo sviluppo di applicazioni pilota valutate da diversi gruppi di stakeholder



Grazie per l'attenzione!



INTERSTAT
Open Statistical Data
Interoperability Framework

cef-interstat.eu
info@cef-interstat.eu



Martino Maggio
INTERSTAT coordinator
martino.maggio@eng.it



Co-financed by the Connecting Europe
Facility of the European Union

The contents of this publication are the sole responsibility of INTERSTAT consortium
and do not necessarily reflect the opinion of the European Union

