

Modello di integrazione in tempo reale del dato statistico con la relativa componente geografica secondo standard europei

Claudio Santoro ^(a), Jody Marca ^(b)

^(a) ISTAT, Dipartimento per la raccolta dati e lo sviluppo di metodi e tecnologie per la produzione e diffusione dell'informazione statistica, Direzione centrale per le tecnologie informatiche e della comunicazione, Viale Oceano Pacifico 171 00144 Roma, tel.: 06-4673.3941, e-mail: claudio.santoro@istat.it

^(b) Collaboratore ISTAT, Viale Oceano Pacifico 171 00144 Roma, e-mail: jmarca@gmail.com

Riassunto

La proiezione geografica delle banche dati delle statistiche territoriali, consente di ampliare la possibilità di fruizione ed analisi dell'informazione stessa.

Il modello, presentato in questo lavoro, risponde a quanto proposto, permettendo l'introduzione della *Location Intelligence* nel pieno rispetto delle linee guida dell'Agenda Digitale Italiana e della più ampia strategia EU2020.

Tale infrastruttura informatica integra in tempo reale il dato statistico mediante protocollo *Statistical Data and Metadata eXchange (SDMX)*, con la componente spaziale per mezzo del formato *Geography Markup Language (GML)* in accordo alla direttiva comunitaria *Infrastructure for Spatial Information in the European Community (INSPIRE)*.

Il sistema in oggetto garantisce inoltre l'industrializzazione del processo produttivo di applicativi *Web e Mobile* atti alla fruizione ed analisi del dato statistico con componente geografica.

Attualmente è disponibile un *Proof of Concept* che permette di navigare in tempo reale il *data warehouse* Istat, in linea con l'attuale sistema di diffusione ufficiale, arricchendo l'informazione statistica con la proiezione territoriale.

Abstract

The geographic view of the data base on territorial statistics enhances data access and fruition, creating new opportunities for the analysis of the information.

The model presented in this contribution allows the introduction of the Location Intelligence in full compliance with the Italian Digital Agenda guidelines and with the wider EU2020 strategy.

This IT infrastructure integrates in real time the statistical data, using the Statistical Data and Metadata eXchange (SDMX) protocol, with the spatial component, through the Geography Markup Language (GML) and in line with the European Commission's Directive Infrastructure for Spatial Information in the European Community (INSPIRE).

The system guarantees the industrialization of the production of web and mobile applications that are functional to the fruition and analysis of the statistical data with a geographic component.

At the moment a Proof of Concept is available. It permits to navigate in real time Istat's data warehouse, in line with the current official diffusion system, and it enhances the statistical information with the territorial projection.

Introduzione

L'Agenda Digitale è una delle sette iniziative principali individuate nella più ampia Strategia EU2020, che punta alla crescita inclusiva, intelligente e sostenibile dell'Unione.

L'Agenda Digitale è stata presentata dalla Commissione Europea nel maggio 2010, ed è stata sottoscritta da tutti gli Stati membri che si sono impegnati a recepirla ed applicarla.

Lo scopo dell'Agenda Digitale è sfruttare al meglio il potenziale delle tecnologie dell'informazione e della comunicazione per favorire l'innovazione, la crescita economica e la competitività.

Nel quadro dell'Agenda Digitale Europea, l'Italia ha elaborato una propria strategia nazionale, individuando priorità e modalità di intervento, nonché le azioni da compiere e da misurare sulla base di specifici indicatori, in linea con gli *scoreboard* dell'Agenda Digitale Europea.

In accordo con EU2020, relativamente la statistica, nell'ambito di ESS Vision 2020 (*European Statistical System*), sono state individuate cinque macro aree (*Key area*). Particolare interesse nella tematica trattata in questo articolo, sono la *Key area* 3, riguardante le nuove sorgenti dati (amministrative, geospaziali, *big data*, etc.), la *Key area* 4, relativa alla definizione di architetture *enterprise* quali *framework* comuni di riferimento e la *Key area* 5 rivolta agli aspetti di strumenti e metodologie di diffusione e visualizzazione.

Nell'ambito di tale contesto, l'Istituto Nazionale di Statistica Italiano, pone come obiettivo strategico quello di diffondere e comunicare in modo efficace l'informazione statistica e le analisi realizzate per favorire la conoscenza della realtà economica, sociale ed ambientale dell'Italia e migliorare i processi decisionali dei soggetti privati e delle istituzioni pubbliche. Per far questo l'Istituto si avvale di diversi strumenti di diffusione, tra cui applicativi *web* che interrogano e navigano le banche dati.

Nell'ambito delle banche dati statistiche territoriali, la proiezione geografica consente di ampliare la possibilità di fruizione ed analisi dell'informazione stessa, dando modo all'utenza di avvalersi di funzionalità intuitive di visualizzazione ed interrogazione che l'introduzione di tale proiezione introduce.

Modello di integrazione

L'enorme mole di informazione statistica costantemente prodotta ed aggiornata, la costante evoluzione amministrativa territoriale, le linee guida europee e nazionali, nonché i principi di *design* architeturale, impongono la non duplicazione delle banche dati, quanto piuttosto l'interrogazione in tempo reale dei sistemi ove tale informazione è contenuta.

Al fine di mantenere disaccoppiata l'informazione statistica da quella geografica, il modello ricorre a standard di *design* architeturale e protocolli affermati per lo scambio dell'informazione in oggetto.

Il sistema è provvisto di una componente software che è in grado di eseguire contemporaneamente delle richieste sia alle banche dati statistiche che ai servizi territoriali elaborando ed integrando le risposte ricevute al fine di facilitare la loro consultazione da parte dell'utente attraverso l'applicazione *web* sviluppata.

Il disaccoppiamento è un obiettivo importantissimo da raggiungere poiché l'inserimento e l'aggiornamento delle banche dati statistiche avvengono con una frequenza tendenzialmente

superiore da quelle territoriali. Le componenti geografiche, come suggerito dalla direttiva comunitaria INSPIRE, risiedono solo dove vengono effettivamente gestite ed aggiornate, questo permette di evitare la duplicazione del dato ed i possibili disallineamenti.

La possibilità di mantenere un forte disaccoppiamento tra dato statistico e geografico permette di avere la possibilità di integrare nuove fonti dati anche gestite da soggetti e pubbliche amministrazioni diverse dall'Istituto Nazionale di Statistica Italiano.

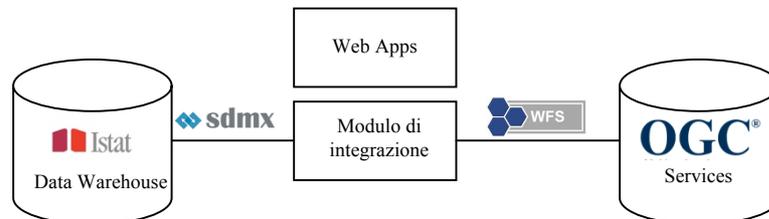


Figura 1 Schema di integrazione tra Data Warehouse dell'Istituto e i servizi geografici

Relativamente la componente statistica, il modello fa uso del protocollo SDMX (*Statistical Data and Metadata eXchange*), standard statistico e tecnico in grado di armonizzare e diffondere dati e metadati statistici in maniera efficiente.

Tale protocollo, sviluppato a partire dal 2001 da sette organizzazioni internazionali (Eurostat, Bank of International Settlements, European Central Bank, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations Statistical Division, World Bank), è oggi lo standard più affermato per lo scambio e la condivisione di dati statistici nella comunità statistica globale e su di esso sono basati molti progetti Stat2015 in Istat. Attualmente l'Istituto Nazionale di Statistica Italiano, si avvale di tale protocollo per la diffusione dei dati del corporate data warehouse I.Stat. L'informazione statistica è pertanto interrogabile anche mediante *Web Service* attraverso un meccanismo *query/download*.

Al contempo la banca dati geografica utilizzata e gestita all'interno dell'Istituto Nazionale di Statistica Italiano deve essere anch'essa fruita attraverso approcci standard. Si è deciso di definire ed implementare dei servizi conformi allo standard OGC *Web Feature Service* il quale permette l'interrogazione delle componenti geografiche e delle informazioni alfanumeriche ad esse associate attraverso il formato standard *Geography Markup Language* (GML) anch'esso standardizzato da OGC. La scelta del formato e del relativo servizio non è stata casuale poiché la loro adozione è aderente ai suggerimenti contenuti nella direttiva INSPIRE oltre ad essere largamente implementato e supportato dai maggiori software GIS sia *Open Source* che proprietari.

Proof of concept

I principi architetturali precedentemente riassunti sono stati implementati durante la realizzazione di un *Proof of concept* che permette la navigazione in tempo reale del data warehouse Istat, in linea con l'attuale sistema di fruizione ufficiale, arricchendo l'informazione statistica con le componenti territoriali.

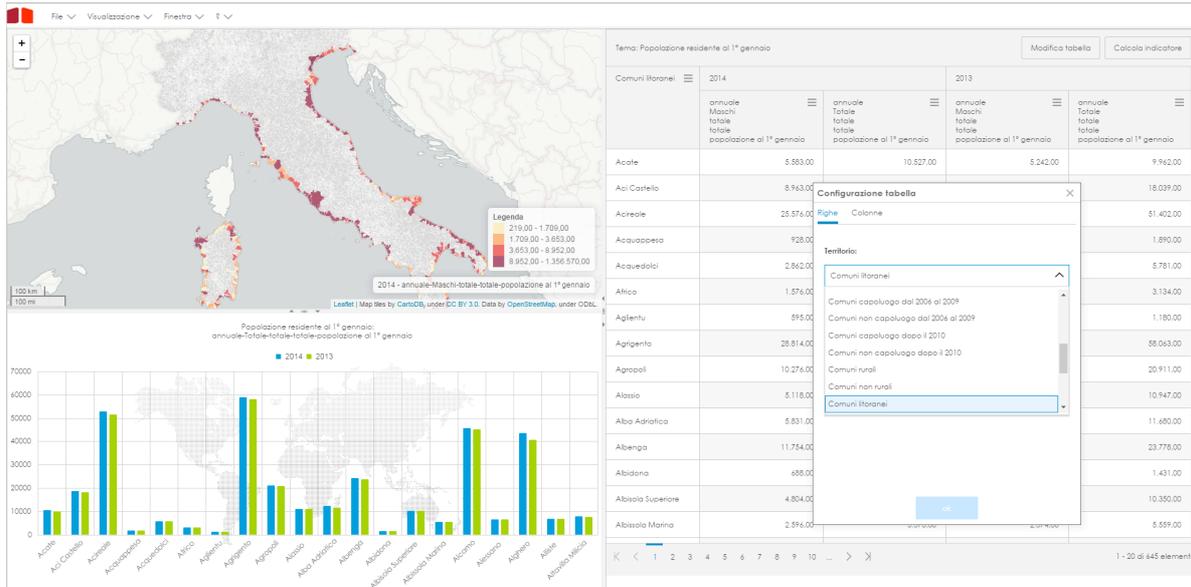


Figura 2 - L'applicativo web prodotto mostra l'informazione richiesta in tre modalità differenti e coordinate tra di loro: rappresentazione tabellare, grafica e cartografica. Ogni area è interattiva e prevede diverse modalità di personalizzazione.

L'informazione statistica è suddivisa per tema e sottoargomenti in conformità con l'attuale sistema di diffusione ufficiale Istat. Oltre al classico approccio di navigazione in uso nell'istituto il sistema è inoltre interrogabile per parola chiave.

I dati sono presentati sotto forma di strutture multidimensionali che gli utenti possono navigare ed esportare in formato tabellare (CSV) e/o grafico (PDF).

L'informazione statistica consultabile attraverso la struttura tabellare è inoltre visibile contemporaneamente anche in modo grafico attraverso la sua proiezione geografica in base al livello territoriale scelto: ripartizioni, regioni, province, comuni (con possibilità di visionare delle sotto suddivisioni quali i soli comuni litoranei, rurali, capoluoghi, livello di urbanizzazione, etc.), SLL (Sistemi locali del lavoro), distretti, etc..

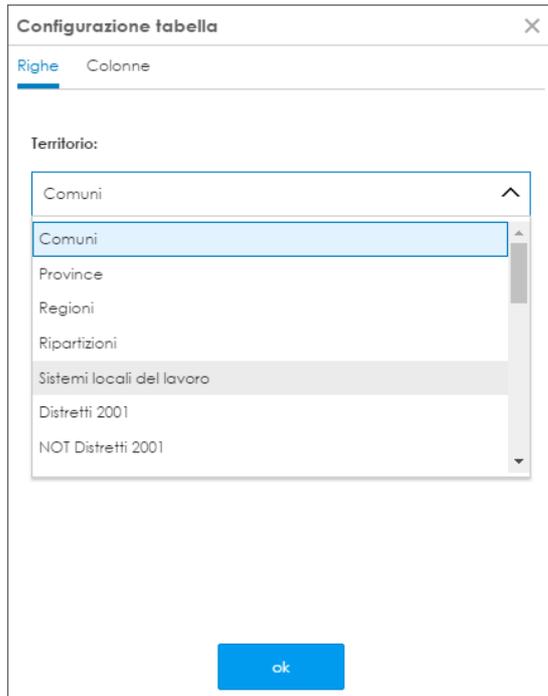


Figura 3 - E' possibile definire il livello amministrativo, sia aggregando (in tempo reale) l'informazione per livelli amministrativi superiori (es: dai comuni alle province, etc.) che per proprietà di un dato livello amministrativo (es: solo i comuni rurali, etc.)

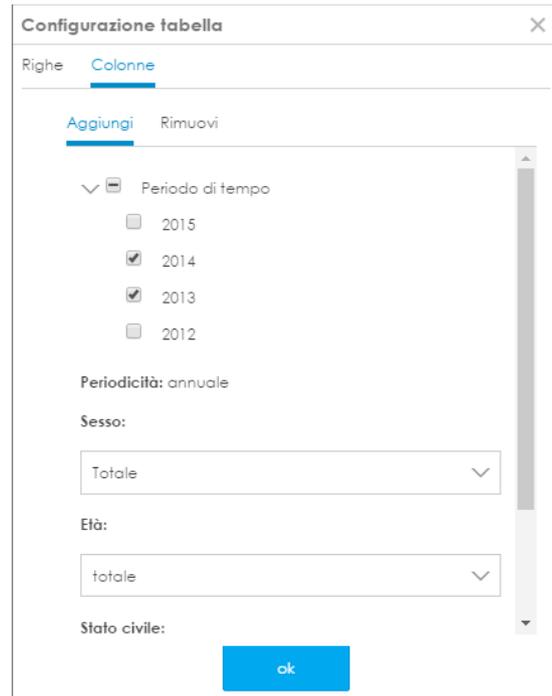


Figura 4 - E' possibile navigare l'informazione n-dimensionale in modo da creare l'aggregazione richiesta. E' inoltre possibile visualizzare la serie storica dell'aggregazione richiesta.

È possibile avere una rappresentazione dei livelli amministrativi superiori attraverso l'aggregazione dei livelli territoriali. Questa soluzione consente ad esempio di calcolare e visualizzare, in tempo reale, le Nomenclature delle unità territoriali statistiche (NUTS) a partire da un livello di dettaglio amministrativo ancora superiore. Ad esempio, nel PoC illustrato, il dato è computato a partire dal livello comunale.

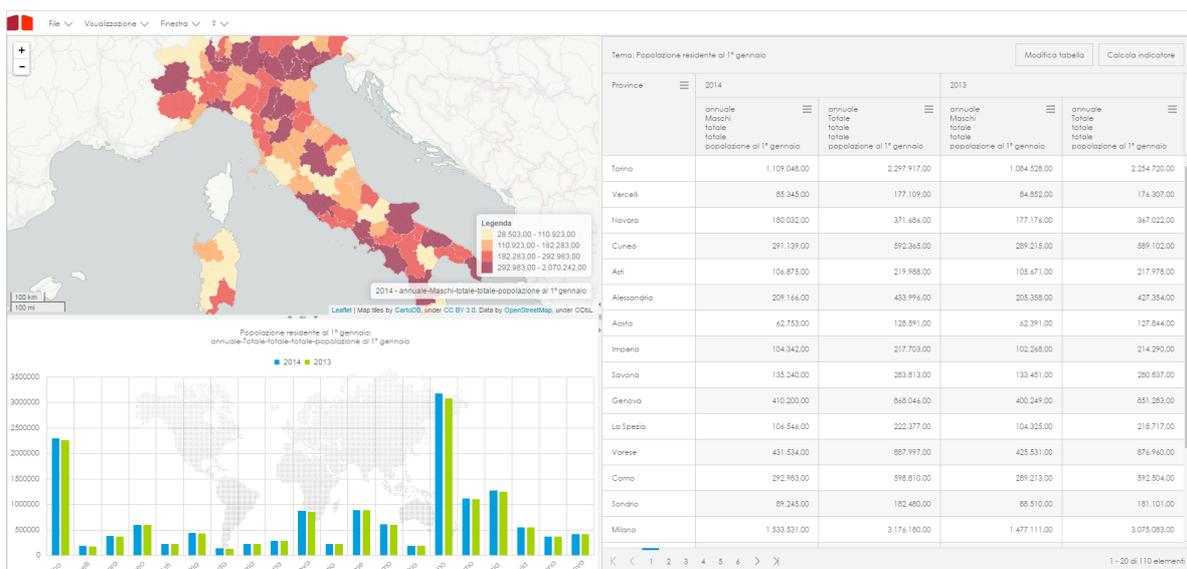


Figura 5 - Esempio di visualizzazione coordinata della serie storica di un'interrogazione definita dall'utente e aggregata in tempo reale a livello provinciale.

La sezione grafica è dinamica: è possibile filtrare le voci rappresentate nel grafico e/o navigare il grafico stesso. E' inoltre possibile scegliere tra differenti sistemi di rappresentazione (istogramma, a settori circolari, etc.), con la possibilità di attribuire dinamicamente ed in tempo reale i valori agli assi.

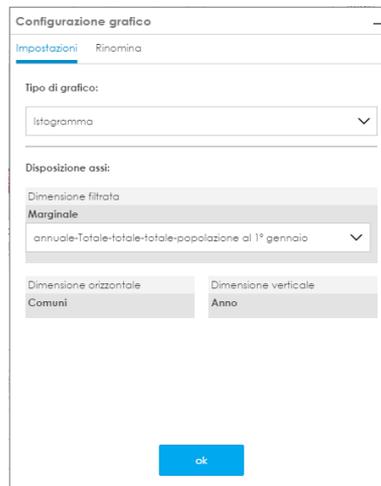


Figura 6 - Finestra di personalizzazione del grafico interattivo. E' possibile: scegliere tra diverse tipologie di rappresentazioni grafiche, disporre sugli assi le variabili e rinominare le variabili stesse al fine di consentire un'esportazione idonea per essere poi riutilizzata.

Il sistema consente di creare, in tempo reale, indicatori mediante l'uso dello strumento "Calcola indicatore". Tali indicatori sono confinati nello spazio utente e possono essere visualizzati in modalità tabellare, grafica e cartografica.

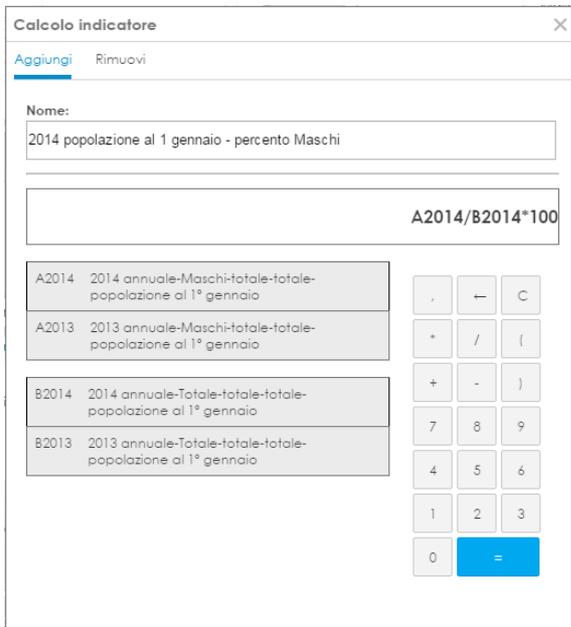


Figura 7 - E' possibile creare, in tempo reale, nuovi indicatori per mezzo della sezione "Calcolo indicatore".

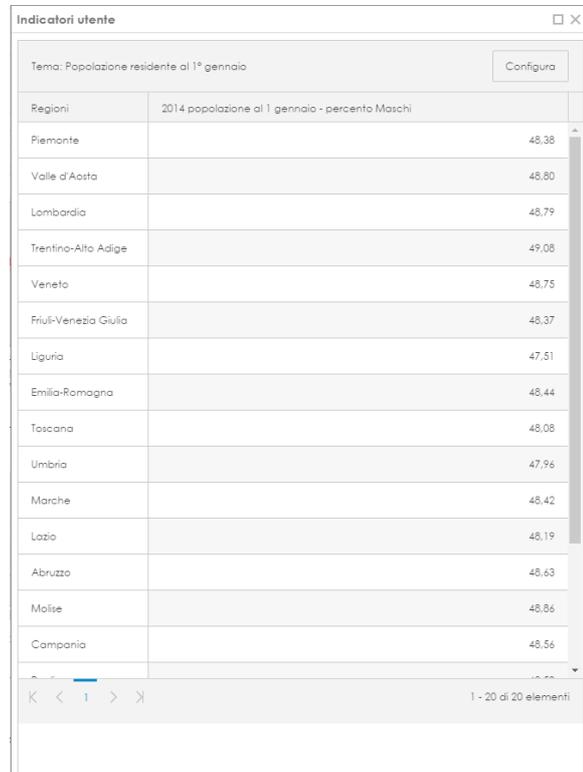


Figura 8 - Gli indicatori utente, frutto dell'esito di "Calcolo indicatori", vengono confinati nello spazio utente diversificando la statistica ufficiale da quanto prodotto dall'utente durante la sua sessione di lavoro.

Conclusioni

L'approccio descritto e la relativa soluzione architeturale sono attualmente impiegate nello sviluppo della linea Atlanti Statistici Istat.

Già nell'ambito del progetto "Informazione statistica territoriale settoriale per le politiche strutturali 2010-2015" finanziato tramite il PON Governance e Assistenza tecnica FESR 2007-2013, è stato prodotto l'applicativo Asti (Atlante statistico territoriale delle infrastrutture) ed è in corso di ultimazione l'applicativo Asc (Atlante statistico dei comuni).

L'obiettivo è quello di produrre gli applicativi *web*, basati sulle banche dati statistiche territoriali, mediante un *framework* che consenta modularità ed ampio riuso dei moduli sviluppati, garantendo tempi di sviluppo molto rapidi, facilitando la manutenzione (*Operation & Maintenance*) e l'evoluzione (*Life Cycle Management*) degli applicativi stessi.

Tra le attività con priorità maggiore, oltre alla completa definizione del *framework* di cui sopra, la *roadmap* evolutiva prevede l'introduzione di moduli di interrogazione ed analisi *GIS*.

Riferimenti bibliografici

Fowler M. (2004), "Inversion of Control Containers and the Dependency Injection pattern", <http://martinfowler.com>

Gamma E., Helm R. Johnson R., Vlissides J., (2005) "Design Patterns: Elements of Reusable Object-Oriented Software", Addison-Wesley

Larman C. (2005). "Applying UML and Patterns: An Introduction To Object-Oriented Analysis And Design And Iterative Development", (3rd Edition), Prentice Hall PTR Upper Saddle River, NJ, USA

Porteneuve C. (2007). "Sviluppare Applicazioni Web 2.0", Apogeo

Hogan P. B. (2010). "HTML5 and CSS3: Develop with Tomorrow's Standards Today (Pragmatic Programmers)", 1st edition, Pragmatic Bookshelf

Flanagan F. (2011). "Javascript: The Definitive Guide", 6th Edition, O'Reilly & Associates Inc.

Haverbeke M. (2011), "Eloquent JavaScript: A Modern Introduction to Programming", No Starch Press

Gulabani S. (2013) "Developing RESTful Web Services with Jersey 2.0", Packt Publishing

Kimball R., Ross M., (2013) "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", 3rd Edition, Wiley

Santoro C., Portuesi S., Scialanca S. (2014) "Sviluppo di applicazioni statistiche attraverso le funzionalità Geoprocessing ArcGIS", Supplemento al n° 2-2014 di GEOmedia

Santoro C. (2015) "Statistiche per le politiche di sviluppo a supporto dei decisori pubblici", Istat, 37-44