

Magellano: il Data Warehouse geo-statistico della Regione Abruzzo

Ufficio Sistema Informativo Statistico-Regione Abruzzo

Dott.ssa Antonella Spinosa, Dott.ssa Tiziana Valentino

Abstract

Da quasi 10 anni il sistema Magellano, interamente progettato e realizzato con SAS dal personale dell'Ufficio Sistema Informativo Statistico della Regione Abruzzo, costituisce il portale attraverso il quale i cittadini hanno accesso alle banche dati regionali.

Fin dall'inizio i dati sono stati presentati in forma tabellare, grafica e con il supporto di carte tematiche per illustrare i dati a valenza geografica.

Nel 2008 è iniziato un progetto con l'obiettivo di far evolvere Magellano sia da un punto di vista tecnologico che funzionale.

Le principali novità sono le seguenti:

- il passaggio alla versione 9 di SAS con l'adozione di strutture multidimensionali più efficienti vista la notevole quantità di dati;
- la revisione completa dell'interfaccia utente;
- l'utilizzo di ArcGis Server come motore geografico;
- l'integrazione completa fra la navigazione multidimensionale di SAS e quella geografica utilizzando ST GeoBridge sviluppato da Sistemi Territoriali.

Il Progetto

Introduzione

Magellano costituisce da molti anni lo strumento attraverso il quale vengono pubblicati dati di diverse tipologie (ambiente, sanità, popolazione, ecc.), provenienti da varie fonti ed elaborati dall'Ufficio Sistema Informativo Statistico della Regione Abruzzo.

Magellano è stato interamente progettato e sviluppato all'interno dell'Ufficio ed i dati sono stati costantemente aggiornati nel tempo fino ad oggi, con un continuo aumento di temi e di dimensione.

La tecnologia utilizzata è stata fin dall'inizio SAS, ed i progettisti del sistema ebbero due idee estremamente innovative, che hanno consentito a Magellano di essere ancora uno strumento efficiente ed efficace:

- efficiente, perché è stato simulato un sistema multidimensionale, pur in assenza di questa componente tecnologica, facendo massiccio uso di dati opportunamente aggregati;
- efficace, perché fin dall'inizio si è fatto ricorso a delle mappe tematiche statiche, prodotte con la PROC GMAP di SAS, in aggiunta ai report ed ai grafici.

Lo scorso anno la Regione ha deciso di riprogettare Magellano, in quanto la tecnologia era ormai matura per fare evolvere le due idee su cui si basava il sistema. E' stato ancora scelto SAS, utilizzando un motore multidimensionale non più simulato e consentendo di ottenere rappresentazioni cartografiche non più statiche ma del tutto sincronizzate con i report.

Attraverso l'integrazione delle due tecnologie GIS e BI, è quindi possibile rappresentare i risultati dell'analisi statistica in un sistema di analisi territoriale e, viceversa, utilizzare funzioni di navigazione territoriale come punto di accesso alle funzioni tipiche dell'ambiente analitico.

L'integrazione delle due componenti quindi consente di avere un quadro completo, che non si limita alla "rappresentazione" dei dati in modalità diverse, ma che consente di sfruttare al meglio tutte le potenzialità di entrambe le metodologie.

Definizione delle fonti dati

I dati elaborati e visualizzati tramite Magellano provengono da diverse fonti quali ad esempio: Istat, ASL, Ministeri ed altri enti pubblici. Alcuni dati restano archiviati come “storici”, altri sono soggetti a periodici aggiornamenti ed integrazioni.

Poiché alcune informazioni contenute nel Data Warehouse sono soggette al segreto statistico, Magellano opera automaticamente una censura quando necessario. Inoltre per ottenere particolari informazioni, non divulgabili come statistiche ma estremamente utili come supporto decisionale, il sistema è dotato di un accesso privilegiato tramite codice che identifica l'utente, in aggiunta all'accesso libero per tutti.

Le banche dati sono raggruppate in 17 “settori” che fanno riferimento alle tematiche definite dall'Istat:

1. Ambiente e territorio
2. Popolazione
3. Sanità e previdenza
4. Cultura
5. Famiglia e società
6. Amministrazione pubblica e servizi sociali
7. Giustizia e sicurezza
8. Conti territoriali
9. Lavoro
10. Prezzi
11. Agricoltura
12. Industria
13. Servizi
14. Commercio estero
15. Censimento agricoltura
16. Censimento popolazione e abitazioni
17. Censimento industria e servizi

Ciascun settore comprende più “argomenti” ad esempio il settore della “Sanità e previdenza” include tra gli altri: la mobilità sanitaria, i ricoveri nelle strutture, le cause di morte, gli infortuni, le pensioni, ecc..

Nuovi argomenti possono essere aggiunti in seguito a ulteriori pubblicazioni da parte delle fonti.

Gli utenti

Il Data Warehouse geo-statistico Magellano si rivolge a più tipologie di utenti allo scopo di offrire un'informazione utile alla conoscenza dei fenomeni economico-sociali e alla loro evoluzione grazie alla comparabilità dei dati.

Gli utenti possono essere classificati come “generico” ed “esperto”.

- l'utente “generico” dispone di una serie di report pre-confezionati che permettono un'analisi globale del fenomeno in esame;
- l'utente “esperto”, come ad esempio può essere una Direzione regionale che necessita uno strumento di supporto decisionale politico-economico, può invece avere una visibilità più ampia delle banche dati presenti ed una maggiore interazione con il sistema.

Creazione dei Data Mart

Il sistema sviluppato si basa su Data Mart tematici, in fase di costruzione per tutti i settori prima citati, costruiti seguendo la metodologia collaudata tipica di un Data Warehouse che prevede i seguenti passi:

1. per prima cosa sono state identificate le fonti dati che alimentano il Data Warehouse, che sono sia interne che esterne alla Regione e costituiscono dei sistemi fra loro indipendenti. Questo significa che nei singoli sistemi sorgente, i dati possono essere codificati in modi fra loro non omogenei. Ad esempio comuni codificati con 5 o 6 cifre, importi espressi in Euro o in Migliaia di Euro, ecc;
2. successivamente sono stati definiti i Data Mart di Magellano;
3. sono state quindi definite le dimensioni di analisi specifiche di un determinato Data Mart (ad esempio i DRG per il settore sanitario, le codifiche ATECO per le imprese, ecc.) e quelle comuni a più Data Mart (ad esempio le dimensioni territoriali e temporali, la codifica del genere, ecc.) e queste ultime sono state rese omogenee.
Alcune dimensioni di analisi sono state strutturate in gerarchie dimensionali, necessarie per definire l'ordine con cui vengono eseguite le operazioni di drill down o roll up, tipiche di un sistema di BI;
4. successivamente sono state definite per ciascun Data Mart le misure di interesse desumibili direttamente dai dati o ricavabili da queste tramite un calcolo. Per ciascuna misura sono stati definiti i domini;
5. ciascun Data Mart è stato modellato utilizzando uno Star Schema, mettendo quindi in relazione fra loro le tabelle dimensionali precedentemente identificate con la tabella dei fatti;
6. si è quindi passati alla fase più delicata e lunga dell'intero processo, ovvero quella di ETL (Extraction-Transformation-Loading) in cui i dati vengono trasportati all'interno dei singoli Data Mart, effettuando su di essi tutte le operazioni di correzione, omogeneizzazione, aggregazione necessarie.
Tutto questo processo viene effettuato utilizzando lo strumento visuale che SAS mette a disposizione, ETL Studio, che consente di tenere sotto controllo tutte le elaborazioni compiute e di poter intervenire su queste nel caso di cambiamenti nelle fonti alimentanti.
7. Infine, una volta creati i Data Mart conformi ai vari Star Schema definiti, vengono create le strutture multidimensionali (cubi) che sono alla base delle elaborazioni di tipo OLAP.

Elaborazioni OLAP

Le elaborazioni di tipo OLAP sono state fatte utilizzando:

- il modulo SAS OLAP Viewer come Client per effettuare la navigazione multidimensionale;
- il modulo ST GeoBridge di Sistemi Territoriali come Client per effettuare la navigazione geografica;
- ArcGis Server di ESRI come Server GIS.

L'architettura del sistema è rappresentata dalla **Fig. 1**.

Le esigenze del progetto erano le seguenti:

- Disporre di due ambienti a diversa complessità per quanto riguarda la parte OLAP: uno semplice per gli utenti pubblici ed uno più complesso per gli utenti esperti;
- Avere un ambiente di navigazione geografica che fosse contestuale con la navigazione OLAP e con un elevato livello di interattività, almeno per gli utenti esperti.

Poiché l'interfaccia utente della versione standard di SAS OLAP Viewer non è modificabile e le funzionalità della componente GIS sono sembrate insufficienti, è stato necessario personalizzare le due interfacce.

E' stato perciò utilizzato il prodotto sviluppato da Sistemi Territoriali, denominato ST GeoBridge, che consiste in un'applicazione web composta da due moduli che interagiscono fra di loro:

- Client OLAP: consiste in un'applicazione web, sviluppata in ambiente Java, che permette la selezione e la visualizzazione dei dati multidimensionale ed è identico al modulo SAS nativo, in quanto utilizza le medesime classi SAS utilizzate da OLAP Viewer.

Questo Client comunica con il Server SAS per la lettura dei cubi e mette a disposizione dell'utente le principali funzionalità di navigazione multidimensionale in modo del tutto identico a SAS OLAP Viewer. Questo è possibile perché sono state utilizzate le medesime classi utilizzate da SAS;

- Client GIS: consiste in un'applicazione web, sviluppata in ambiente .NET, per la visualizzazione di mappe tematiche, che contiene anche un modulo di integrazione con l'ambiente SAS. Questo Client comunica con il Server GIS tramite dei web services (Web services GIS) e consente la visualizzazione delle mappe tematiche mettendo a disposizione dell'utente le principali funzionalità di navigazione GIS;
- un modulo di amministrazione dei metadati, il Metadata Manager, che consente la definizione dei metadati necessari per l'integrazione dei due ambienti. Esso consente di definire i campi da usare sia nei report SAS, sia nei layers ESRI, per poter realizzare la tematizzazione delle mappe, così come i metadati necessari per la selezione del dato da visualizzare.

L'utente del sistema ha a disposizione le seguenti funzionalità:

- **Selezione del dato da visualizzare.**

L'utente seleziona il dato da visualizzare a partire da un elenco predefinito organizzato per ambito tematico.

- **Visualizzazione dati multidimensionali.**

Dopo aver selezionato i dati che si vogliono analizzare, si accede all'applicazione per la visualizzazione dei dati multidimensionali che mostra il risultato con un layout predefinito. In questa finestra è possibile modificare il layout del report selezionando le dimensioni, le variabili, filtri, ecc..

L'utente ha a disposizione le funzionalità basiche di navigazione in ambiente multidimensionale:

- Definizione del layout: questa funzionalità permette di impostare le dimensioni e le variabili da visualizzare per riga e colonna;
- Drill-down (i bottoni al fianco dell'elemento di una gerarchia permettono di espandere /scendere di livello nella gerarchia);
- Espand;
- Down;
- Impostazioni di filtri sulle dimensioni.

La visualizzazione tabellare permette anche di navigare i dati visualizzati tramite menù e strumenti. Selezionando il nome di una dimensione in una gerarchia compare un menù che permette di realizzare operazioni di:

- Navigazione (espansione);
- Ordinamento dei dati;
- Ruota tabella (scambia righe con colonne);
- Esportazione di dati in Excell;
- Evidenziazione;
- Filtri;
- Totali;
- Ecc..

Diversi tipi di grafico possono essere attivati tramite una barra degli strumenti o tramite il menù "visualizza".

Quando il report contiene una dimensione definita, nella meta informazione, come geografica, il cubo visualizzato è collegato ad una mappa (servizio di ArcGis Server) e viene attivato un apposito bottone che indica la possibilità di ottenere una rappresentazione dei dati in una mappa tematica. La mappa viene tematizzata con la prima variabile (misura)

visualizzata nel report ed è sincronizzata con il report multidimensionale, nel senso che tutte le operazioni realizzate nel report sono realizzate anche sulla mappa e viceversa.

Sulla mappa possono essere realizzate sia operazioni cartografiche (zoom, pan, ecc.) sia operazioni di navigazione dei dati (drill down, ecc.) :

- Navigazione della mappa;
- Attivazione e disattivazione di un tema;
- Visualizzazione delle informazioni relative ad elementi geografici individuati con il mouse;
- Impostazione della scala di visualizzazione;
- Selezione di un oggetto su tema attivo;
- Selezione mediante rettangolo;
- Tematizzazione:
 - Impostazione del metodo di tematizzazione;
 - Modifica dei colori della tematizzazione;
 - Scelta della variabile da tematizzare;
 - Cambio del livello geografico della mappa (drill down geografico).

Il sistema realizza la sincronizzazione fra la componente SAS e la componente ESRI. Cambiando il livello di dettaglio territoriale di un cubo SAS si aggiorna la mappa tematica e, viceversa, cambiando il livello di dettaglio territoriale della mappa tematica, si aggiorna la visualizzazione del cubo SAS. Inoltre, i filtri applicati in ambiente SAS sono applicati anche in ambiente geografico. In questo modo le due viste dei dati sono sincronizzate in qualsiasi momento.

Presentiamo un esempio di integrazione tra la navigazione geografica e il dato tabellare dell'elaborazione OLAP per il settore Sanità della regione Abruzzo.

Le figure sono relative alla Mobilità Sanitaria e mostrano il numero dei ricoveri che si effettuano in Abruzzo di residenti in altre regioni (**Fig. 2**) e di residenti nelle province della regione Lazio (**Fig. 3**).

Conclusioni

Il progetto rappresenta una evoluzione tecnologica e funzionale molto importante rispetto alla versione attuale di Magellano, e vuole essere sempre più uno strumento sia di conoscenza che di supporto dei processi decisionali.

Riferimenti Autori

Antonella Spinosa
Tel. 0862-363618
Fax. 0862-363215
e-mail: antonella.spinosa@regione.abruzzo.it
Ufficio Sistema Informativo Statistico - Regione Abruzzo
Via Leonardo Da Vinci n.6
67100 L'Aquila

Tiziana Valentino
Tel. 0862-363656
Fax. 0862-363215
e-mail: tiziana.valentino@regione.abruzzo.it
Ufficio Sistema Informativo Statistico - Regione Abruzzo
Via Leonardo Da Vinci n.6
67100 L'Aquila

Bibliografia

- Catitti A.C., Cococcia L. (2001) "Statistics web warehousing", Convegno SAS European User Group - Firenze
- Catitti A.C., Cococcia L. (2001) "StatiztiKai Adattàrhàz a Web-en", Convegno SAS Group ungherese - Budapest
- Catitti A.C. (2008) "L'integrazione della statistica con la cartografia", Convegno SAS forum - Milano
- Catitti A.C. (2008) "Cartografia + Business Intelligence = Conoscenza", Rivista ITA.SAS.COM n° 5
- Catitti A.C. (2008) "Statistica e territorio", Rivista INNOV@ZIONE.PA , Anno V- N.11

Ambito del Progetto

Il progetto si inserisce nell'ambito delle attività svolte dall'Ufficio Sistema Informativo Statistico della Regione Abruzzo, volte al perseguimento della conoscenza e della comprensione dei principali fatti riguardanti la realtà regionale.

Tempi e Fasi di Realizzazione

La fase di realizzazione del progetto ha avuto inizio lo scorso novembre partendo dalla tematica della sanità considerata strategica e certamente la più complessa per la mole di dati da trattare.

Per quanto riguarda il "settore" sanità i primi Data Mart previsti per la pubblicazione saranno la mobilità sanitaria e i motivi di ricovero nelle strutture abruzzesi, mentre, sono previsti per la popolazione la struttura e il movimento anagrafico. Tutti gli altri verranno pubblicati in modo progressivo man mano che verranno completati.

Software Utilizzati:

Tipo	Software	Produttore
Server GIS	ArcGIS 9	ESRI
BI	SAS 9	SAS Institute
BI & GIS	ST Geo Bridge	Sistemi Territoriali s.r.l.

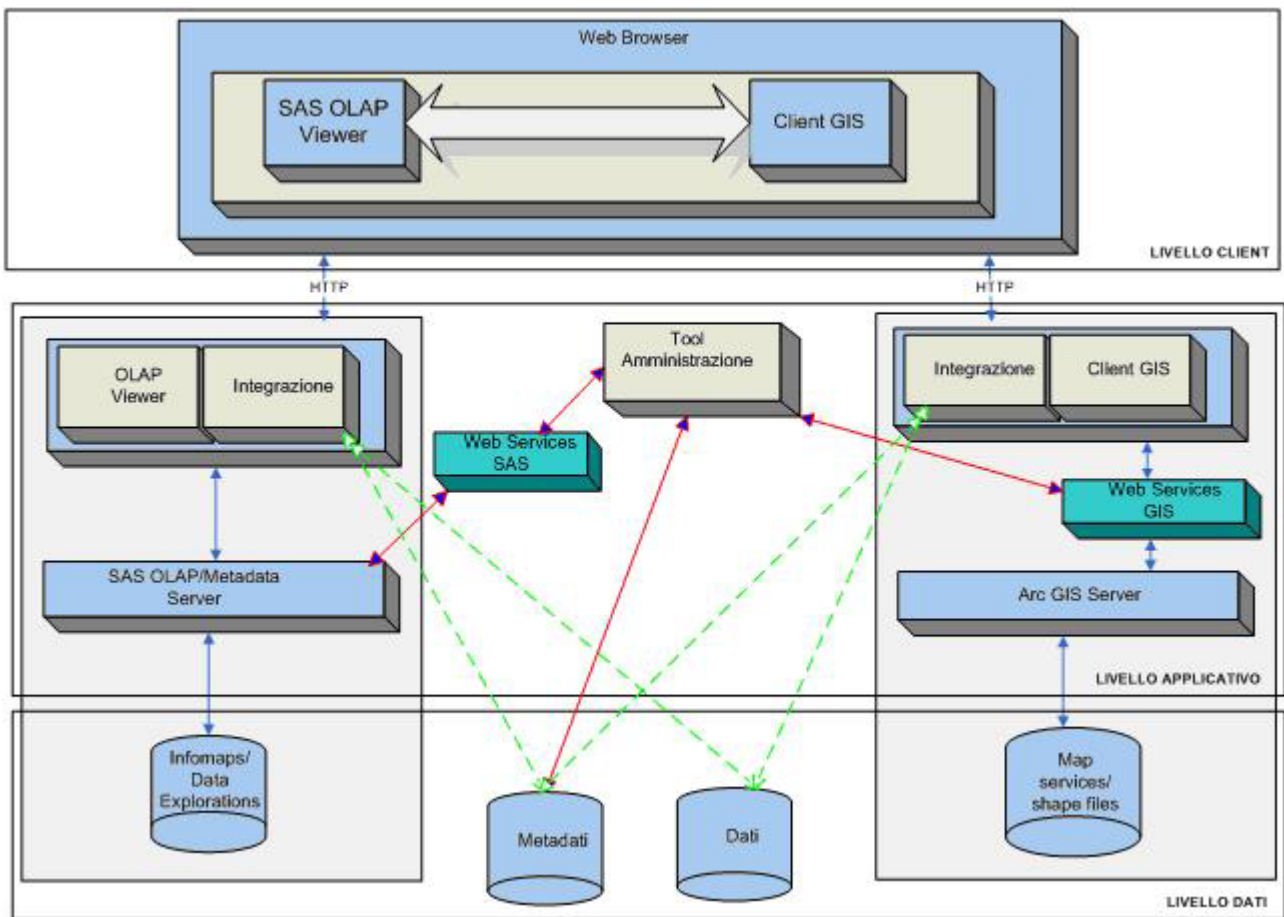


Fig. 1 - L'architettura del sistema.

Percorso di navigazione: Filtri applicati:

Visualizza▼ Dati▼ Proprietà Colonne 1 - 5 di 19

Regione	Basilicata	Calabria	Campania	Emilia Romagna	Friuli Venezia Giulia
	Somma Di Num Ric	Somma Di Num Ric	Somma Di Num Ric	Somma Di Num Ric	Somma Di Num Ric
ANNO					
2004	268	429	1,707	475	79
2005	265	435	1,703	490	66
2006	354	419	1,661	495	73



Fig. 2 - Mobilità Sanitaria relativa al numero dei ricoveri effettuati in Abruzzo di residenti in altre regioni.

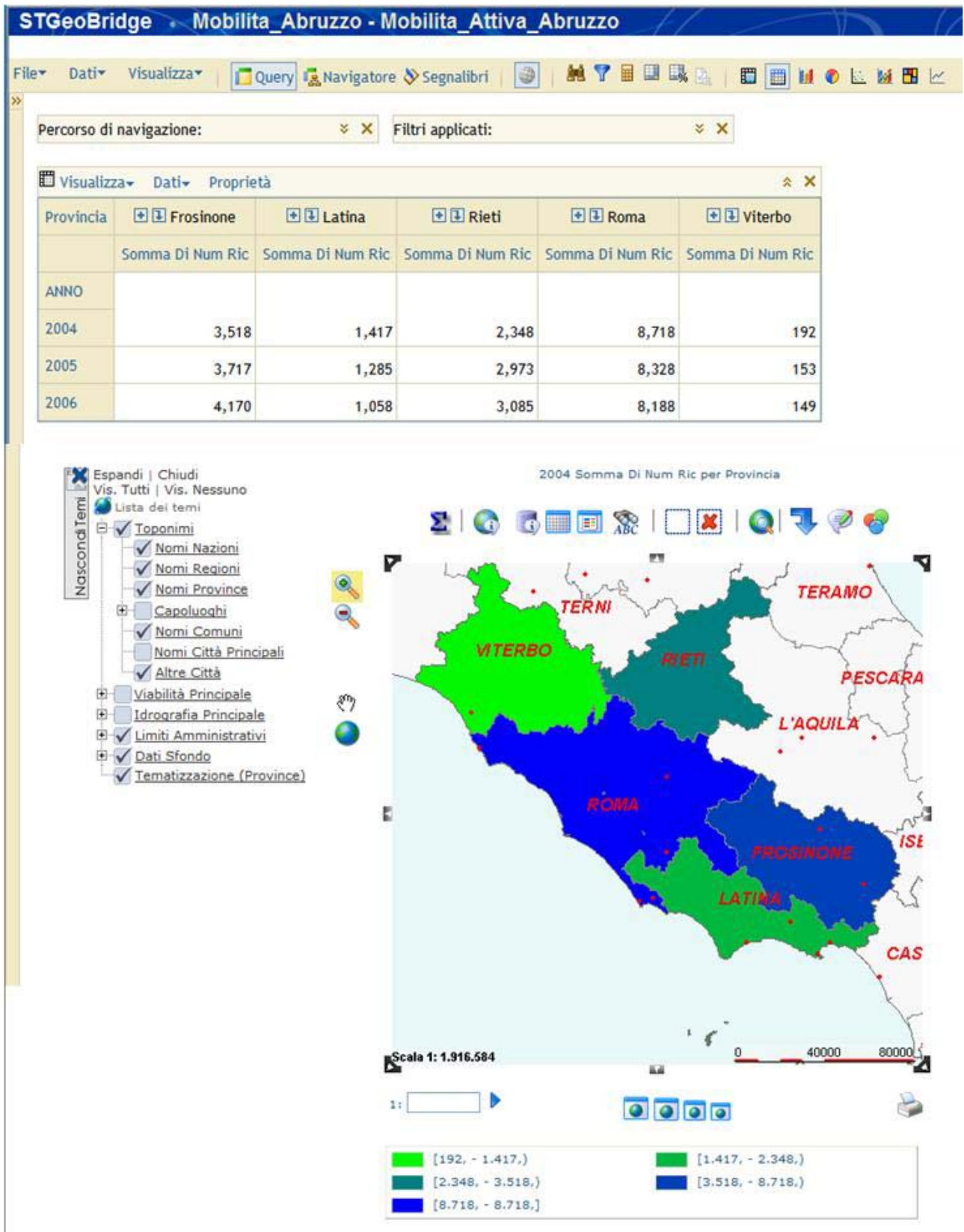


Fig. 3 - Mobilità Sanitaria relativa al numero dei ricoveri effettuati in Abruzzo di residenti nelle province della regione Lazio.