

Sesta conferenza nazionale di statistica  
Roma, 6-8 novembre 2002

## I Sistemi Informativi Geografici e la qualità del dato statistico

Fabio Crescenzi  
Istat, Via A. Ravà 150 00142 Roma, [crescenz@istat.it](mailto:crescenz@istat.it)

### Sintesi

L'esigenza di informazioni sempre più ricche e dettagliate sulla realtà sociale ed economica spinge verso l'impiego di tecniche nuove, capaci di agevolare la produzione tempestiva di dati di qualità a tutti i livelli di dettaglio territoriale, anche per piccole aree. I Sistemi Informativi Geografici (*GIS*) consentono oggi una gestione più razionale ed oculata dei dati aumentando la possibilità di accedere ad informazione riferita ad aree di interesse definite in modo anche personalizzato. Questa nuova disponibilità può comportare rischi anche considerevoli e questi, se non tenuti adeguatamente sotto controllo, possono avere conseguenze rilevanti del rispetto degli standard di qualità del dato statistico.

**Parole chiave:** Sistemi Informativi Geografici (*GIS*), Address Matching

### 1. Introduzione

Nel presente lavoro si cercherà di analizzare quali sono le caratteristiche di alcuni di questi strumenti con l'obiettivo da un lato di proporre nuove modalità di impiego, ma anche, da un altro punto di vista, di individuare le determinanti principali, capaci di migliorare la affidabilità dei dati statistici riferiti al territorio. Questo con riferimento particolare alle informazioni tipicamente contenute in un Sistema Informativo Geografico considerando l'utilizzo dell'insieme di tutte le informazioni ausiliarie sul territorio che sono raccolte nel corso delle operazioni censuarie e che nell'ambito di un Sistema Informativo Geografico possono essere gestite.

E' ben noto che la quantità e la qualità di informazione statistica varia molto secondo il dettaglio territoriale al quale si vuole condurre l'analisi. A livello regionale o provinciale, l'informazione è abbastanza ricca e aggiornata perché sono disponibili dati di fonte amministrativa, dati delle rilevazioni correnti, dati delle indagini campionarie. Scendendo a livello comunale o sub-comunale, la disponibilità dei dati statistici si riduce drasticamente e la fonte di riferimento principale è costituita dai censimenti che, se pur ricchi, tendono ad invecchiare rapidamente. I censimenti hanno cadenza decennale e l'informazione da loro desumibile diventa ben presto datata; parte dei fenomeni rilevati con i censimenti sono strutturali e non si modificano rapidamente,

anche questi tuttavia a livello micro-territoriale possono essere oggetto di modificazioni anche molto rapide.

## **2. Le caratteristiche principali di un Sistema Informativo Geografico. Le potenzialità nella gestione, produzione e diffusione di informazione statistica**

Le applicazioni dei *GIS* a problemi di carattere statistico sono sempre più numerose ed evolute. Da una fase iniziale nella quale il *GIS* era visto e utilizzato come strumento esclusivo di rappresentazione visiva, ci si sta muovendo rapidamente verso un sempre più ampio impiego del *GIS* come strumento completo di analisi spaziale. La statistica applicata ai diversi settori di indagine, partendo da quello ambientale, nei quali il *GIS* mostra con evidenza le sue potenzialità, per arrivare alle tradizionali indagini socio-demografiche, non ha più come obiettivo esclusivo quello di proporre dati e loro più o meno efficaci visualizzazioni, ma anche e soprattutto quello di proporre e consentire la costruzione di modelli interpretativi, di proporre stime che devono però essere accompagnate da misure della loro affidabilità.

L'impiego dei *GIS* nelle indagini statistiche è d'altra parte sempre più diffuso sia con riferimento a progetti aventi rilevanza internazionale che con riferimento a progetti sviluppati in ambito nazionale e locale. La disponibilità sia per gli utenti pubblici che privati dei dati geografici e statistici relativi al Censimento della Popolazione e delle Abitazioni del 1991, che sono ancora oggi molto richiesti, ha rappresentato una occasione per far proliferare e diffondere applicazioni geostatistiche per diversi usi: dal supporto alla pianificazione territoriale, alle analisi di geomarketing.

L'applicazione dei Sistemi Informativi Geografici nei vari settori applicativi è d'altra parte strettamente legata alla disponibilità di dati geografici e statistici. Il mercato dei dati geografici è oggi ancora interessato da molteplici problemi ma vede anche importanti opportunità: se da un lato continua a persistere una sostanziale inefficienza nella formazione dei prezzi, nelle modalità di cessione e nei formati di realizzazione, dall'altro lato progetti importanti come quello del nuovo Sistema Cartografico di riferimento, lasciano intravedere nuovi e importanti sviluppi.

Un *GIS* è in sostanza un insieme complesso di componenti hardware, applicazioni software, per acquisire, processare, analizzare, immagazzinare, restituire in forma grafica o alfanumerica dati riferiti al territorio. In un *GIS* trovano rappresentazione oggetti opportunamente georeferenziati, posti in relazione con tabelle di attributi, di tipo geografico, alfanumerico o di tipo statistico. Con la costruzione di un *GIS* ci si pone l'obiettivo di rispondere a quesiti relativi a dove gli oggetti sono collocati e quali caratteristiche hanno in determinati ambiti territoriali. Funzioni di tipo geometrico, quali la misura di distanze, di aree, o la ricerca di centroidi, della distanza più breve fra un punto ed una linea, sono facilmente gestibili con tali software. Queste funzioni possono essere estremamente importanti in analisi statistiche territoriali, come nel caso della determinazione di aree di centri abitati o delle distanze relative agli spostamenti casa-lavoro, ecc.

Una tipologia di applicazioni molto utili riguarda la determinazione delle distanze che separano unità territoriali e di conseguenza le unità statistiche che sono in esse comprese. Come noto la distanza euclidea non soddisfa quasi mai le esigenze analitiche:

occorre determinare i percorsi minimi che si ottengono considerando la rete stradale esistente, la presenza di collegamenti ferroviari, aerei, etc. Tali problemi possono essere più facilmente risolti sfruttando le funzioni di ottimizzazione dei percorsi (*routing*) che si basano sugli stessi algoritmi utilizzati nella *car navigation* e le funzioni *GIS* applicate ai grafi stradali, ferroviari, etc. Tali funzioni richiedono un punto di partenza, un punto di arrivo, eventuali punti di sosta, nonché il parametro di ottimizzazione che si vuole utilizzare (la distanza, il tempo, il traffico o anche una combinazione di più parametri). Questa, come altre applicazioni sono esaminate in Biallo (2002).

Il termine *GIS* è usato in modi diversi in contesti diversi: in taluni casi si fa riferimento a sistemi hardware e software per il trattamento di dati spaziali, in altri casi a specifiche categorie di prodotti software, in altri ancora ad archivi a base geografica e ad applicazioni informatiche costruite per la loro gestione. Più di recente si è affermato il riferimento ai sistemi *GIS* come approccio complessivo e logica di trattamento dei dati più che ad uno strumento specifico, sia esso sistema informativo o strumento software.

Più in dettaglio si possono esaminare quali sono gli elementi caratterizzanti un *GIS*. In primo luogo un sistema *GIS* consente la integrazione di informazioni sovrapponendo strati informativi diversi. In ognuno di questi strati le codifiche e gli attributi possono essere riferite ad elementi geometrici di natura diversa che possono essere punti, linee o poligoni. Questo consente di integrare in un unico contenitore informazioni che possono riguardare unità statistiche collocabili puntualmente, oppure unità statistiche riferibili ad elementi lineari o poligonali. Gli elementi poligonali possono essere di diversa forma e ampiezza. In conseguenza di ciò è possibile procedere alla integrazione ed al confronto di dati riferibili ad oggetti collocati o collocabili puntualmente attraverso coordinate geografiche, con dati riferibili solamente ad elementi lineari come strade o ad aree.

Un insieme di funzionalità che caratterizza un *GIS* rispetto ad altri sistemi, è costituito dalle cosiddette funzioni di “overlay” o di sovrapposizione di oggetti geografici. Queste permettono di generare una nuova mappa cartografica che sintetizza elementi geografici ed attributi ricavati dai dati di partenza.

Una delle applicazioni più interessanti ed utili dal punto di vista statistico è proprio quella della stima di dati per una mappatura (detta mappatura obiettivo o target) in aree diverse da quelle di raccolta o disponibilità dei dati. Le tecniche di overlay producono conseguenze che devono essere correttamente gestite sia dal punto di vista geometrico che dal punto di vista statistico. Con un *GIS* si ottiene la mappatura intersezione fra la mappatura di origine e quella obiettivo. Dal punto di vista geometrico uno dei problemi più rilevanti riguarda la gestione dei micropoligoni che frequentemente sono originati nella intersezione delle due mappature, da un punto di vista statistico deve invece essere considerato in modo particolare il problema dei modelli che sono soggiacenti alle stime che si producono per le aree obiettivo.

Alcuni prodotti *GIS* consentono di ricostruire l'informazione sulla base di predefiniti insiemi di regole. Si deve però considerare che i modelli soggiacenti possono non essere adeguatamente soddisfacenti da un punto di vista statistico e che non sempre la qualità del dato statistico che si ottiene è adeguatamente misurata e documentata. Uno dei principali aspetti sui quali occorre intensificare la ricerca è proprio quello della misura della qualità e della documentazione relativa ai dati statistici.

Con riferimento più generale alle principali misure della qualità di un *GIS* ricordiamo gli usuali parametri quali la accuratezza posizionale, la accuratezza tematica, la

accuratezza temporale, la congruenza logica, la completezza che possono essere misurati attraverso indicatori di qualità (si veda Biallo, 2002, pag.65-70).

### **3. Le nuove metodologie di stima del dato territoriale e il ruolo cruciale dei metadati**

Una sfida, non di poco conto, nel matrimonio fra geografia e statistica riguarda lo sviluppo di nuove applicazioni della metodologia statistica capaci di sfruttare opportunamente l'integrazione dei dati statistici con i dati geografici e gli strumenti di tipo "GIS" (una interessante rassegna su questi aspetti è contenuta in Openshaw, 1995). Un esempio di questo tipo riguarda la costruzione di modelli per la stima dei dati quando l'informazione è disponibile per aree geografiche non completamente coincidenti con quelle alle quali siamo interessati. Queste metodologie sono chiamate tecniche di interpolazione areale (*Areal Interpolation Techniques*) e sono più facilmente implementabili sfruttando le funzioni di overlay fornite nei software GIS.

Una descrizione delle principali tecniche di interpolazione areale è contenuta in Flowerdew et. al. (1991), Goodchild et al. (1993) and Fischer and Langford (1995) e un riepilogo è dato in UNITED NATIONS (2000, pag.108).

Questi metodi distinguono le zone di origine (*source zones*) per le quali l'informazione è disponibile dalle zone obiettivo (*target zones*) per le quali si vuole ricostruire l'informazione.

Il metodo più appropriato da utilizzare dipende dalle ipotesi che è possibile assumere: la distribuzione uniforme nelle zone di origine, nelle zone obiettivo, oppure in una terza categoria di zone dette zone di controllo (*control zones*).

Il panorama delle metodologie di interpolazione areale può essere notevolmente ampliato includendo tecniche che consentono di sfruttare variabili correlate con le variabili di interesse e mediante le quali sia possibile stimare l'informazione per le aree di interesse. Non ci addentreremo oltre nell'esame di queste metodologie rinviando alla bibliografia citata e più in generale alla vasta letteratura presente su questa materia. Deve tuttavia essere chiaro che questi metodi sono affetti da un errore che può divenire anche molto forte nel caso di applicazioni che richiedano grande accuratezza. E' altrettanto evidente però che l'interpolazione areale può rappresentare l'unica possibilità per ottenere informazione per piccole aree quando non sia possibile percorrere strade alternative, come quella delle stime mediante le usuali tecniche campionarie per gli elevati costi che necessiterebbero, o la possibilità di ottenere informazione affidabile per somma esatta da aree di più piccola dimensione, o ancora da fonte amministrativa.

Anche in questo caso diviene perciò essenziale la documentazione associata all'informazione, fornendo i metadati relativi alle metodologie di produzione accompagnate possibilmente da valutazioni dell'errore che potrebbe essere conseguente al loro utilizzo.

#### **4. La qualità del disegno delle basi territoriali, primo requisito per la produzione di una informazione statistica territoriale affidabile<sup>1</sup>**

Nel 1991, con il progetto CENSUS, per la prima volta le basi territoriali, ovvero i limiti amministrativi, delle località e delle sezioni di censimento, sono stati correttamente inquadrati dal punto di vista geografico, acquisiti e resi disponibili in formato digitale.

Da strumento esclusivo di organizzazione della raccolta dati, le basi territoriali si sono così evolute e rappresentano oggi uno strumento essenziale per la analisi del territorio consentendo di sfruttare le potenzialità dell'integrazione dei dati statistici con dati geografici mediante le nuove, più potenti, e facilmente gestibili, applicazioni informatiche.

CENSUS 2000 è stato realizzato utilizzando molteplici fonti informative, fra le quali spiccano le ortofoto aeree in formato digitale messe a disposizione dall'Agea (ex AIMA), le Carte Tecniche Regionali e Comunali, i grafi con il disegno delle strade, ferrovie, idrografia in formato digitale vettoriale. Attraverso queste, si è proceduto ad aggiornare i limiti di località ed anche di sezione, migliorandone la qualità geografica e fornendo un prodotto completo ai comuni, lasciando comunque ad essi, che sulla base del regolamento anagrafico sono i detentori delle basi, il compito di controllo e validazione. Le basi territoriali CENSUS 2000 sono in fase di revisione e saranno completate e rese disponibili per la diffusione nel corso del 2003.

L'obiettivo che ci si è posti con la costruzione della base territoriale non, è stato quello di costruire cartografia in senso stretto, ma è stato funzionale all'esigenza di disporre di uno strato tematico per la raccolta e la associazione al territorio del dato statistico. Il risultato è tanto più soddisfacente, quanto più può basarsi su fonti cartografiche di qualità e di riferimento comune per l'utenza. Per questo motivo l'Istat è stato particolarmente attento alla integrazione con i sistemi informativi territoriali e le cartografie principalmente utilizzate a livello locale.

L'obiettivo di rendere quanto più possibile compatibile la cartografia derivata dal censimento con le cartografie prodotte localmente, è raggiunto utilizzando, ove disponibili, queste ultime come fonte primaria nell'acquisizione di elementi di delimitazione di località e sezioni di censimento, e con le ortofoto che rappresentano uno strato di verifica e aggiornamento e di copertura delle aree nelle quali tali cartografie non siano disponibili.

Per ottenere una adeguata omogeneità territoriale una prima importante distinzione è fatta fra le aree urbane, le aree adibite prevalentemente ad attività industriali o commerciali, e le altre aree. Per questo motivo il primo passaggio prevede l'identificazione delle località abitate (centri e nuclei abitati), e delle località produttive dal territorio comunale residuo (case sparse). Le località produttive sono introdotte per la prima volta in questa tornata censuaria. Con la loro introduzione non ci si propone di comprendere e delimitare tutte le aree produttive. Esistono infatti situazioni, in particolare nei centri abitati, ove abitazioni, stabilimenti industriali e centri commerciali sono presenti in modo misto e la delimitazione delle relative aree risulta assai complessa. Il disegno delle località produttive riguarda solo realtà presenti sul territorio extraurbano e che rispondano a prefissati limiti dimensionali.

---

<sup>1</sup> A cura di Fabio Lipizzi

La delimitazione delle sezioni di censimento è stata effettuata con dei criteri differenziati per il territorio urbano ed extraurbano.

Il criterio generale è quello di far corrispondere i relativi limiti ad evidenti elementi “fisici” come strade, ferrovie, corsi d’acqua, ecc, al fine di dare ai rilevatori degli elementi certi per riconoscere la zona loro assegnata. Eccezioni sono contemplate solo in quei casi nei quali con questo criterio non sia possibile “chiudere” la sezione entro dimensioni ragionevoli o in quei casi nei quali non sia garantita la possibilità di ricostruire zone amministrative o di interesse statistico.

Nei grandi centri, per i quali si dispone dei grafi stradali, e comunque nei comuni con popolazione superiore ai 30.000 abitanti, l’obiettivo è di far corrispondere la sezione di censimento con il singolo isolato.

La costruzione delle sezioni di censimento 2000 si propone come riferimento comune di tutti i censimenti, compreso quello dell’agricoltura. Nei centri urbani ha seguito la stessa logica del 1991, anche se con una migliore qualità geometrica. Nelle aree extraurbane si è proceduto invece ad una revisione per consentire la georeferenziazione delle aziende agricole. La soluzione individuata ha richiesto che le sezioni di censimento fossero disegnate rispondendo all’esigenza di corrispondere a somme di fogli di mappa, nel rispetto comunque dei criteri generali di identificabilità fisica delle stesse. Un altro aspetto rilevante, è la riduzione della superficie territoriale delle attuali sezioni di censimento extraurbane, in media circa 522 ettari, ad una estensione media inferiore ai 400 ettari. Tale circostanza ha fatto lievitare il numero delle sezioni di censimento extraurbane di oltre un terzo dalle attuali 54.416 e di conseguenza il numero complessivo delle sezioni di censimento dovrebbe superare nel conteggio finale le 400.000.

L’analisi della distribuzione della popolazione per tipologia di sezione di censimento mostra che la popolazione, si trova per il 91% localizzata nei centri mentre nei nuclei e nelle case sparse vive solamente il 3% e 6% rispettivamente. La relativa estensione in termini di superfici territoriali mostra invece che centri e nuclei coprono circa il 10% del territorio e che le “case sparse” incidono invece per valori prossimi al 90%, pari a 28,5 milioni di ettari circa.

Le sezioni di censimento sono concepite in modo da permettere la ricostruzione per somma, oltre che di tutte le località abitate, anche dei principali aggregati territoriali sub-comunali. L’Istat ha inoltre ricostruito in concerto con i comuni interessati la suddivisione in zone sub-comunali intermedie (circoscrizioni, quartieri, ecc.). Tale ricostruzione è già disponibile per i comuni a più grande dimensione ed interessa un numero sempre più ampio di comuni.

Verrà predisposto l’aggiornamento dell’archivio degli “itinerari di sezione”, ovvero delle porzioni delle strade (dette archi di strada) presenti nella sezione cui è associato il nome della strada e dei numeri civici estremi che cadono nella stessa. L’informazione per aggiornare l’archivio degli itinerari di sezione è stata derivata in modo parziale dagli archivi 1991 e in parte dai grafi stradali, per la parte restante è fornita dai comuni cui spetta comunque la validazione complessiva.

Nella Figura 1 è mostrato un esempio di disegno delle sezioni extraurbane in un’area campione della provincia di Verona. Per disegnare le sezioni di censimento extraurbane è utilizzata la base per l’interpretazione morfologica del terreno (in questo caso l’ortofoto dell’Agea), sono oscurate le località abitate e produttive (nella figura in nero), e avendo come riferimento le delimitazioni dei fogli di mappa catastali, si è proceduto al

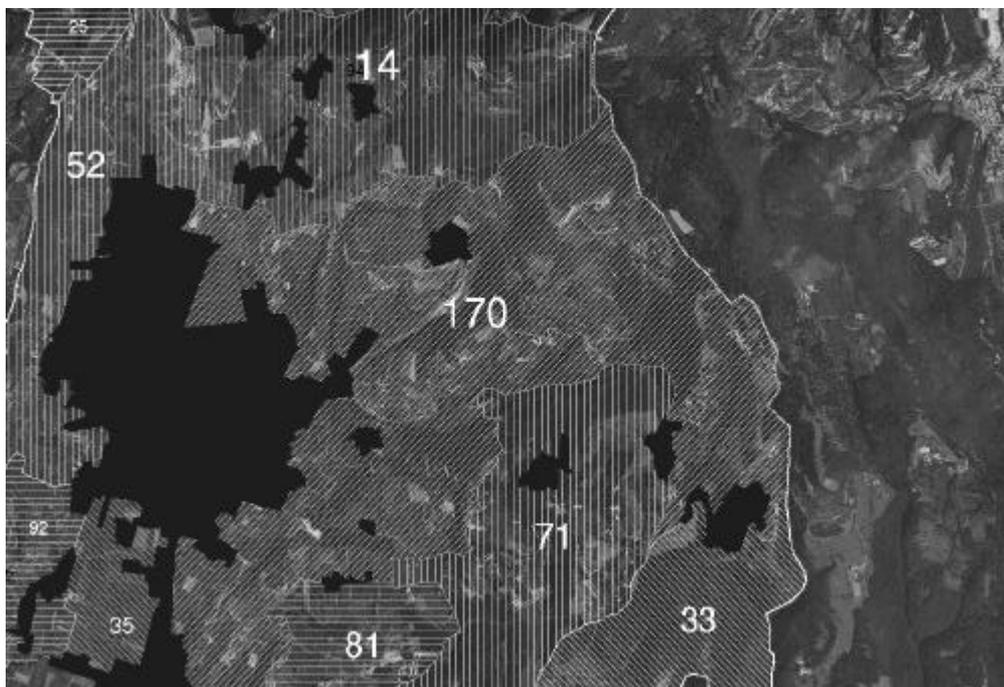
loro accorpamento in sezioni (nella figura rappresentate dai poligoni retinati) nel vincolo della delimitazione delle stesse con strade, fiumi, etc.

**Tabella 1:** *Numero di sezioni di censimento, popolazione, superficie e densità di popolazione per località abitata (dati censimento 1991)*

<b>CENTRI ABITATI</b>				
	Sezioni (numero)	Popolazione	Superficie <sup>2</sup> (ha)	Densità
Nord	104.983	22.783.084	780.395	29,19
Centro	46.556	9.587.018	316.406	30,30
Sud e Isole	78.360	19.047.500	437.532	43,53
<b>Italia</b>	<b>229.899</b>	<b>51.417.602</b>	<b>1.534.333</b>	<b>33,51</b>
<b>NUCLEI ABITATI</b>				
	Sezioni (numero)	Popolazione	Superficie (ha)	Densità
Nord	23.543	872.691	79.106	11,03
Centro	7.774	356.286	32.400	11,00
Sud e Isole	7.870	444.960	34.007	13,08
<b>Italia</b>	<b>39.187</b>	<b>1.673.937</b>	<b>145.513</b>	<b>11,50</b>
<b>CASE SPARSE</b>				
	Sezioni (numero)	Popolazione	Superficie (ha)	Densità
Nord	30.306	1.673.419	11.125.508	0,15
Centro	11.706	968.049	5.487.073	0,18
Sud e Isole	12.404	1.045.024	11.832.441	0,09
<b>Italia</b>	<b>54.416</b>	<b>3.686.492</b>	<b>28.445.022</b>	<b>0,13</b>
<b>ITALIA PER RIPARTIZIONI GEOGRAFICHE</b>				
	Sezioni (numero)	Popolazione	Superficie (ha)	Densità
Nord	158.832	25.329.194	11.985.009	2,11
Centro	66.036	10.911.353	5.835.879	1,87
Sud e Isole	98.634	20.537.484	12.303.980	1,67
<b>Italia</b>	<b>323.502</b>	<b>56.778.031</b>	<b>30.124.868</b>	<b>1,88</b>

<sup>2</sup> Stimata da GIS

**Figura 1** *Disegno delle sezioni extraurbane in un'area campione della provincia di Verona.*



##### **5. Le codifiche delle unità territoriali nella loro evoluzione temporale e l'interconnessione dei sistemi mediante l'individuazione di definizioni, classificazioni, standards comuni**

Le fonti alle quali poter attingere informazione statistica a livello sub-comunale sono di tre tipi: fonti di tipo amministrativo, e quindi rintracciabili all'interno delle singole amministrazioni locali territoriali; fonti che provengono da rilevazioni dirette, siano esse totali o campionarie; fonti che derivano le informazioni da rilevazioni ufficiali o semi-ufficiali, anch'esse totali o campionarie. La disponibilità di informazione statistica territoriale potrà crescere se si sarà capaci di far crescere le possibilità di interconnessione e di dialogo tra i diversi archivi esistenti introducendo definizioni, classificazioni e standard comuni.

Ad ogni dato può essere associata una codifica relativa alla collocazione temporale o spaziale dell'unità. Una componente che diviene sempre più importante nella valutazione dell'errore è proprio quella dell'errore di codifica spazio/temporale del dato. Particolare attenzione deve essere posta alla identificazione e alla classificazione delle suddivisioni del territorio di tipo amministrativo che rivestono importanza ai fini statistici, alla armonizzazione delle nomenclature e gestione delle variazioni, ed al potenziamento della accessibilità ad una informazione condivisa.

Come noto le principali funzioni amministrative a livello locale sono svolte da Regioni, Province e Comuni. La legge 142/90 riconosce come Ente locale anche le Comunità

montane e nell'art. 17 modificato nella legge 265/99, le aree metropolitane, anche se l'iter per la definizione di quest'ultime risulta oggi non completato. Tutti questi enti hanno come elemento costitutivo il territorio. Esistono poi altre suddivisioni a carattere amministrativo che rivestono interesse per la statistica, sono le suddivisioni in circoscrizioni, quartieri, rioni, ed altre aree sub-comunali, e le suddivisioni cui fanno riferimento le principali amministrazioni dello Stato ed enti di interesse pubblico così come fotografate alla data del 1997 nel Primo atlante di geografia amministrativa, prodotto dall'Istat.

Le informazioni sui comuni definiti montani e parzialmente montani, di quelli facenti parte delle Comunità Montane, la rispettiva popolazione e superficie, montana e non, sono state alla base della realizzazione dell'Atlante statistico della Montagna Italiana (Istat 1999), e l'aggiornamento di questi dati è una attività che potrà rendere possibile la realizzazione di nuovi prodotti per la diffusione.

In generale, occorre identificare tutte le suddivisioni che risultano di interesse strategico per la statistica, gestire le codifiche, le variazioni, migliorare l'accessibilità a queste informazioni nel Sistan. Le attività dovranno proporsi l'obiettivo di realizzare un sistema sempre più efficiente ed integrato per la gestione di informazioni ricche e dettagliate sui principali contesti territoriali utilizzati per l'analisi della realtà sociale e produttiva.

Le suddivisioni a carattere amministrativo possono essere desunte da fonti amministrative, anche se in molti casi si pone il problema di dare corretta applicazione e in modo condiviso delle norme vigenti. Le codifiche di tali unità territoriali rappresentano uno strumento essenziale per il riferimento delle statistiche al territorio. Con l'aggiornamento di tale elenco, secondo le modifiche che le unità stesse subiscono, ci si propone di garantire la confrontabilità nel tempo, relativamente al livello territoriale scelto, dei dati ottenuti.

Per tale motivo, gli archivi informativi sulle unità amministrative, quali sono oggi Regione, Provincia e Comune, e le variazioni territoriali e di nome che le hanno modificate dal 1861 al 2000, costituiscono un patrimonio storico di notevole interesse, oltre che per l'Istat ed il SISTAN, per l'attività di molte Amministrazioni pubbliche, degli Enti locali, nonché dei ricercatori e del mondo accademico. Su tali informazioni è stato realizzato un volume con tutte le variazioni storiche dal 1861 al 2000: "Unità amministrative. Variazioni territoriali e di nome dal 1961 al 2000".

Nel sistema informativo storico delle amministrazioni territoriali (SISTAT) sono documentati:

- le validità temporali delle Unità Amministrative (U.A.) contemplate dal sistema
- le variazioni territoriali (cessione o annessione di zone, costituzione o soppressione di U.A.)
- le variazioni di nome;
- la composizione delle U.A. di livello gerarchico superiore
- le principali caratteristiche territoriali e strutturali (superficie e popolazione legale)

Una particolare funzione di interrogazione permette di ricostruire l'ambito territoriale di una U.A. che ha subito nel tempo variazioni. In tal modo è possibile valutare correttamente i confronti temporali degli aggregati associati alle U.A., evitando la distorsione dovuta alla differente composizione territoriale.

Le unità amministrative presenti sul territorio italiano, così come le conosciamo oggi, sono il risultato di adattamenti a spinte di diversa natura. La normativa vigente, ad

esempio, contempla la possibilità di costituire un nuovo Comune, previa consultazione delle popolazioni interessate. Diverso è invece il caso delle annessioni e delle cessioni di territorio verso paesi esteri in conseguenza dei conflitti bellici, che nel passato hanno interessato il nostro Paese.

Le informazioni di base disponibili sono rappresentate dalle caratteristiche delle unità amministrative, dai relativi codici, dalla loro esistenza o meno in un certo periodo, dalle pertinenze territoriali, dalle relazioni che nel tempo si sono stabilite tra unità contigue in seguito al verificarsi di fenomeni di soppressione e successiva aggregazione ad unità confinanti, al distacco di territori da una unità amministrativa e conseguentemente alla costituzione di nuove entità o loro aggregazione a quelle contigue. I fenomeni si fanno ancora più complessi se a questo si aggiunge l'esistenza di più livelli amministrativo-territoriali, le cui relazioni sono definite in modo gerarchico. Ne segue che anche la modifica di una unità amministrativa ad un certo livello determina la conseguente variazione del livello inferiore, come nel caso di modifica di confini per la costituzione di nuove Province e la relativa variazione di dipendenza amministrativa dei comuni interessati.

Le informazioni associate al territorio interessato dalla variazione sono, inoltre, la superficie territoriale e la popolazione legale, ossia la popolazione che, al censimento precedente la variazione occorsa, risultava esservi residente. L'aggiornamento di tale dato risulta di fondamentale importanza, anche per il corretto confronto dei dati censuari.

## **6. I sistemi di riconoscimento mediante l'address matching e gli effetti della qualità della toponomastica**

L'utilizzo di archivi è lo strumento più efficace per poter disporre di informazione che sia territorialmente molto dettagliata ed aggiornata con continuità. Tuttavia di solito negli archivi non sono contenute informazioni dettagliate sulla collocazione territoriale di una unità statistica. Più frequentemente è disponibile l'indirizzo e il nome o codice del comune.

L'Istat si è posto il problema di tenere aggiornate le informazioni statistiche riferite alla griglia delle sezioni di censimento e quindi la possibilità di "ancorare" sul territorio anche informazioni non rilevate con il censimento.

L'utilizzo di strumenti *GIS* rende possibile la acquisizione e gestione dell'informazione su unità statistiche collocate in modo puntuale qualora siano note le coordinate precise in cui questa si trova. Questo in genere non avviene e più di frequente si conosce l'indirizzo.

L'*address matching*, letteralmente riconoscimento di indirizzi, è un sistema per la geocodifica dei dati a partire dagli indirizzi. Con esso è possibile collocare un punto su una mappa non tramite coordinate classiche (ad es. latitudine e longitudine) ma tramite indirizzo (via, numero civico, ecc.).

L'Istat ha realizzato *SISTER*, un sistema basato sull'*address matching* per la gestione e la analisi dell'informazione statistica a livello micro-territoriale, anche derivata da fonte non censuaria, purchè dotata di indirizzo.

Derivando l'informazione da un modello ausiliario di censimento, per ciascuna sezione di censimento, già a seguito del censimento del 1991 sono stati memorizzati tutti gli archi di strada, ed dei relativi numeri civici, che contornano o attraversano la sezione

stessa. Questo database sarà aggiornato con le informazioni rilevate nel censimento del 2001. Nella Figura 2 è mostrato un esempio di itinerario di sezione 2001 con la rappresentazione su ortofoto.

**Figura 2** Esempio di itinerario di sezione e della sua rappresentazione su ortofoto



Questo archivio permette di descrivere ogni sezione di censimento con l'elencazione degli archi di strada che le appartengono. L'archivio può anche essere usato per la geocodifica: fornendo come *input* al sistema un indirizzo, o una serie di indirizzi, si ottengono in *output* gli indirizzi in forma normalizzata arricchiti dell'attributo "sezione di censimento" di appartenenza.

Il sistema di geocodifica permette di riferire alla griglia delle sezioni di censimento qualsiasi informazione che abbia come attributo un indirizzo. Questo sistema di geocodifica integrato con il Sistema Informativo Geografico offre importanti prospettive per le analisi sul territorio.

Di ogni sezione di censimento sono disponibili infatti le coordinate del centroide che consente, il georiferimento puntuale, anche se approssimato, di unità statistiche per le quali sia stata individuata la sezione di censimento di appartenenza.

Due sono le possibili forme di utilizzo del collocamento puntuale delle unità statistiche sul territorio: la prima riguarda la già menzionata sintesi del fenomeno attraverso la rappresentazione diretta su mappa dei punti relativi alle unità statistiche che possiedono caratteristiche di interesse e l'analisi dei relativi addensamenti.

La seconda forma riguarda la possibilità di consentire all'utente di definire interattivamente aree e poligoni e potere ottenere riepiloghi statistici sulle aree così definite. Una procedura del software seleziona in sostanza le unità statistiche che cadono nell'area selezionata e fornisce statistiche di riepilogo su queste.

Alcuni paesi come ad esempio il Canada hanno realizzato prodotti del genere che sono oggi in grado di consentire di rispondere ad interrogazioni via Internet, anche se ovviamente questo richiede idonee procedure di protezione dal rischio di violazione della riservatezza che sono allo studio (si veda a tal proposito *UN/ECE 1998* e *Vliegen, Van Leeuwen 1998*).

Questo potrebbe in prospettiva fornire una risposta al problema dell'unità areale modificabile (Openshaw, 1985), problema derivante dal vincolare le possibilità di aggregazione delle unità statistiche ad aree definite con dubbia significatività statistica e geografica. Spesso infatti la distorsione che si introduce vincolando la possibilità di analisi ad aree amministrative o comunque ad aree prive di uno specifico significato può risultare anche di notevole portata.

Le performance delle metodologie basate sull'*address matching* dipendono dalla qualità delle informazioni sulle strade presenti negli archivi, comprendendo con tale definizione tutte le tipologie di aree di circolazione esistenti, quindi sia le strade nazionali, che le strade provinciali e le strade comunali, etc.

Anche se le finalità per le quali gli archivi sono realizzati ed i contenuti sono spesso diversi, alcune esigenze possono essere considerate comuni. In particolare è evidente l'importanza della correttezza e della affidabilità delle informazioni sulle strade. Allo stato attuale il panorama si presenta particolarmente diversificato: alcuni archivi possiedono caratteristiche interessanti dal punto di vista dell'aggiornamento, ma sono carenti con riferimento alla correttezza delle denominazioni, altri viceversa contengono informazioni corrette ma non aggiornate.

Sono molti i problemi che riguardano la toponomastica:

- I canali ufficiali di comunicazione delle variazioni sono molto spesso lenti, qualche volta disattesi, frequentemente non disponibili su un idoneo supporto informatico.
- I soggetti che attingono dagli enti proprietari per acquisire informazioni sulle strade sono molti, le relative richieste sono assai difformi, gli standards, informatici e non, sono disomogenei. Conseguentemente gli enti proprietari sono soggetti ad un eccessivo carico di richieste, non sono arricchiti dalla restituzione di basi dati da essi utilizzabili e possono verificare le informazioni circolanti con molta difficoltà.
- Non esiste chiarezza su cosa deve essere accessibile pubblicamente e cosa no.
- Le abbreviazioni, il diverso ordine di presentazione delle singole parole, le denominazioni di uso comune diverse dalle denominazioni ufficiali, rendono molto spesso non riconoscibili gli indirizzi ed assai difficoltoso l'utilizzo incrociato di archivi pubblici.
- Non esistono standards comuni sulla presentazione di un indirizzo, né esiste una codifica o una metodologia unificata per la costruzione di un codice identificativo delle strade.

Uno dei campi nei quali occorrerà procedere in modo molto incisivo è proprio quello della ricerca di standard nella definizione degli indirizzi, nonché della gestione degli archivi delle denominazioni ufficiali e trasmissione delle informazioni sugli aggiornamenti. Uno dei problemi fondamentali nel confronto di informazioni che contengono indirizzi è proprio quello che spesso la qualità con la quale gli indirizzi sono presenti negli archivi è cattiva con denominazioni di strade che spesso si discostano dalle denominazioni ufficiali (si veda anche una analisi analoga svolta nel contesto americano da R. ORLI et al., 1996). Questo provoca una considerevole caduta nel tasso di riconoscimento degli indirizzi e di conseguenza della percentuale di unità statistiche correttamente riferibili al territorio. Recentemente L'AIPA e l'Istat hanno avviato una ricerca comune proprio su questi temi ritenuti di cruciale interesse.

## 7. Address matching: un esempio concreto <sup>3</sup>

Le unità statistiche sotto analisi hanno come attributo l'indirizzo: si tratta dell'archivio dei dipendenti dell'Istituto nazionale di statistica impiegati nelle sedi di Roma. Di questi si vuole studiare la distribuzione sul territorio, per analizzare il fenomeno dello spostamento casa-lavoro ed eventualmente la localizzazione ottimale della sede di lavoro unica.

Per analizzare questi dati con riferimento al territorio, i passi logici sono i seguenti:

1. l'archivio dei dati viene sottoposto alla procedura di "riconoscimento indirizzi e geocodifica" e in tal modo alla fine del processo ciascuna istanza del *file* viene arricchita con le informazioni territoriali archiviate nel sistema (codice della sezione di censimento e coordinate metriche di un punto interno alla stessa). Vedi Figura 3.
2. i risultati ottenuti vengono processati in ambiente GIS: la distribuzione dei punti che corrisponde ai dipendenti viene rappresentata sullo spazio corrispondente, definito tramite il sistema metrico di riferimento utilizzato. La congruità di questi dati con i dati GIS riferiti agli ambiti territoriali amministrativo-censuari (comuni e sezioni di censimento) permette la sovrapposizione dei diversi *layer* (vedi Figura 4).

**Figura 3** Esempio di riconoscimento indirizzi

indimp	locimp	capimp	sigla_prov	cod_sezion	coor_x	coor_y	viauff
VIA MANFREDI EUSTACHIO N. 8	ROMA	197	RM	635032	788960	4647418	VIA MANFREDI EUSTACCHIO
VIA DI VALLERANO N.101 PAL.08 INT.6	ROMA	143	RM	4602003	788048	4631307	VIA VALLERANO
VIA MONTE CERVIALTO, 82 SC. B INT. 16	ROMA	139	RM	3818048	792121	4650764	VIA MONTE CERVIALTO
LARGO PEDIVIGLIANO N.9 (MORENA)	ROMA	40	RM	4480035	800107	4636464	LARGO PEDIVIGLIANO
P.ZZA LOTARIO N. 8 SC. B INT. 2	ROMA	162	RM	713039	792038	4648616	PIAZZA LOTARIO
VIA PIAN DUE TORRI N. 25	ROMA	146	RM	1596162	787336	4639199	VIA DI PIAN DUE TORRI
VIA L. GADOLA, S.N.C. PAL. G INT.8	ROMA	155	RM	5744082	797546	4643121	VIA GADOLA LUIGI
VIA GREGORIO XI, 171	ROMA	169	RM	3680663	783489	4644048	VIA GREGORIO XI
VIA ROBERTO VISIANI N. 127 SC. S INT. 8	ROMA	143	RM	5787016	791598	4636500	VIA VISIANI ROBERTO
VIA DE CALVI GIOVANNI, N. 61	ROMA	151	RM	1780089	786121	4641057	VIA DE CALVI GIOVANNI
VIA DEGLI ABETI, 46 SC. B INT. 16/A	ROMA	172	RM	2489070	796070	4643317	VIA DEGLI ABETI
VIA TUSCOLANA N. 859 SC. B INT. 2	ROMA	174	RM	2939059	795602	4640380	VIA TUSCOLANA
V.LE RONCHI N. 20	ROMA	177	RM	2778085	794605	4644330	VIALE RONCHI
VIA CHIANCIANO, N. 11	ROMA	189	RM	3485004	786727	4651545	VIA CHIANCIANO
V.LE DELLE PROVINCE, 101 SCALA B INT. 15	ROMA	162	RM	5654145	792162	4645771	VIALE DELLE PROVINCE
VIA CASATI GAETANO, N. 6 INT. 04	ROMA	154	RM	1394047	789398	4641171	VIA CASATI GAETANO
VIA DELL'ANNUNZIATELLA, 119 INT.11	ROMA	147	RM	2580054	791161	4639797	VIA DELL' ANNUNZIATELLA
VIA SEBINO N.29 SC.V INT.6	ROMA	199	RM	2351117	790762	4647355	VIA SEBINO
VIA LEONCAVALLO RUGG. N. 3 SC. B	ROMA	199	RM	2252010	791263	4648847	VIA LEONCAVALLO RUGGERO
VIA ADA NEGRI, N.54 INT.9	ROMA	137	RM	3208085	795336	4650275	VIA NEGRI ADA
VIA PISINO N.155 SC. B INT.23 FAB.6	ROMA	177	RM	5738102	795340	4644844	VIA PISINO
VIA CONCA D'ORO N. 246 SC. C INT. 20	ROMA	141	RM	2154067	791895	4649660	VIA CONCA D'ORO
VIA DI VIGNA MURATA N.306 SC. B INT. 3	ROMA	143	RM	3314082	791386	4636612	VIA DI VIGNA MURATA
VIA CAMILLO SBARBARO, 5	ROMA	143	RM	4592023	788960	4634190	VIA SBARBARO CAMILLO
CIRCONV. APPIA, 69 SC. B INT. 8	ROMA	179	RM	1273113	791761	4642129	CIRCONVALLAZIONE APPIA
VIA LANZA GIOVANNI, N. 150	ROMA	184	RM	25083	790023	4644223	VIA LANZA GIOVANNI
VIA COLLALTO SABINO, 124 INT. 2	ROMA	199	RM	2271051	791795	4648388	VIA COLLALTO SABINO
VIA FONTANAROSA N. 25 INT. 24	ROMA	177	RM	952186	794599	4644100	VIA FONTANAROSA
VIA GIOVANNI BUCCO, 61	ROMA	156	RM	3253046	793991	4648895	VIA BUCCO GIOVANNI
VIA GIOVANNI DEVOTI, 28/A GIA' 30 INT.5	ROMA	167	RM	1799104	785062	4645220	VIA DEVOTI GIOVANNI
VIA ALGARDI ALESSANDRO N. 15 INT. 7	ROMA	152	RM	1625175	788904	4643071	VIA ALGARDI ALESSANDRO
VIA IVANDE BONOMI N.74 SC. E INT.6	ROMA	139	RM	3807022	792055	4650468	VIA BONOMI IVANOE
V. BADOERO GIOVANNI ANDREA, 67	ROMA	154	RM	1383052	789320	4641040	VIA BADOERO GIOVANNI ANDREA
VIA FOSSATO DI VICO SC. B N. 10 INT. 08	ROMA	181	RM	1048108	793248	4642523	VIA FOSSATO DI VICO

<sup>3</sup> A cura di Orietta Gargano

Nella Figura 4 sono stati rappresentati:

- le sezioni di censimento del comune di Roma (grigliato nero irregolare)
- i dipendenti Istat, tramite un punto posizionato nel centroide della sezione di censimento alla quale sono stati associati tramite l'indirizzo di residenza
- le sedi dell'Istat di Roma (gli asterischi neri)

La figura è stata fissata dopo un ingrandimento sulla zona centrale della città.

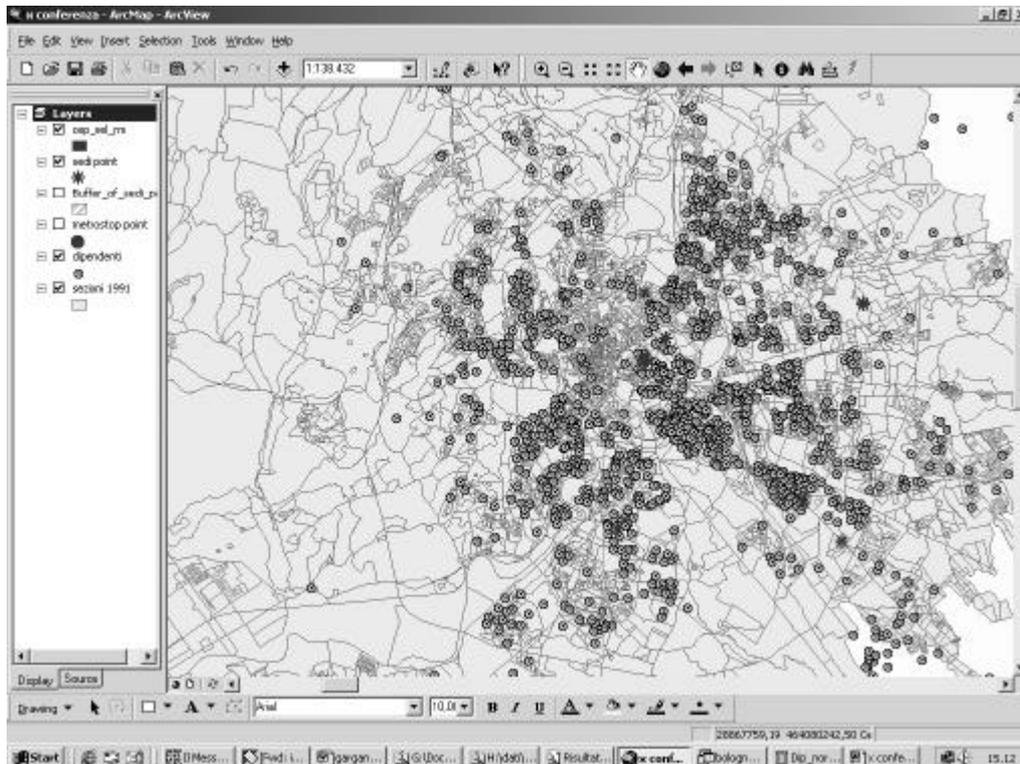
A questo punto è possibile utilizzare le funzioni di analisi spaziale offerte dal GIS per

1. individuare le zone di rispetto (*buffer*) di un punto obiettivo (esempio le sedi Istat) e calcolare quanti individui ricadono in un intorno dato. Nella Figura 5 sono evidenziati i dipendenti che distano meno di tre chilometri dalla sede Istat più vicina
2. calcolare le distanze euclidee di ciascun dipendente dalla propria e/o da tutte le sedi

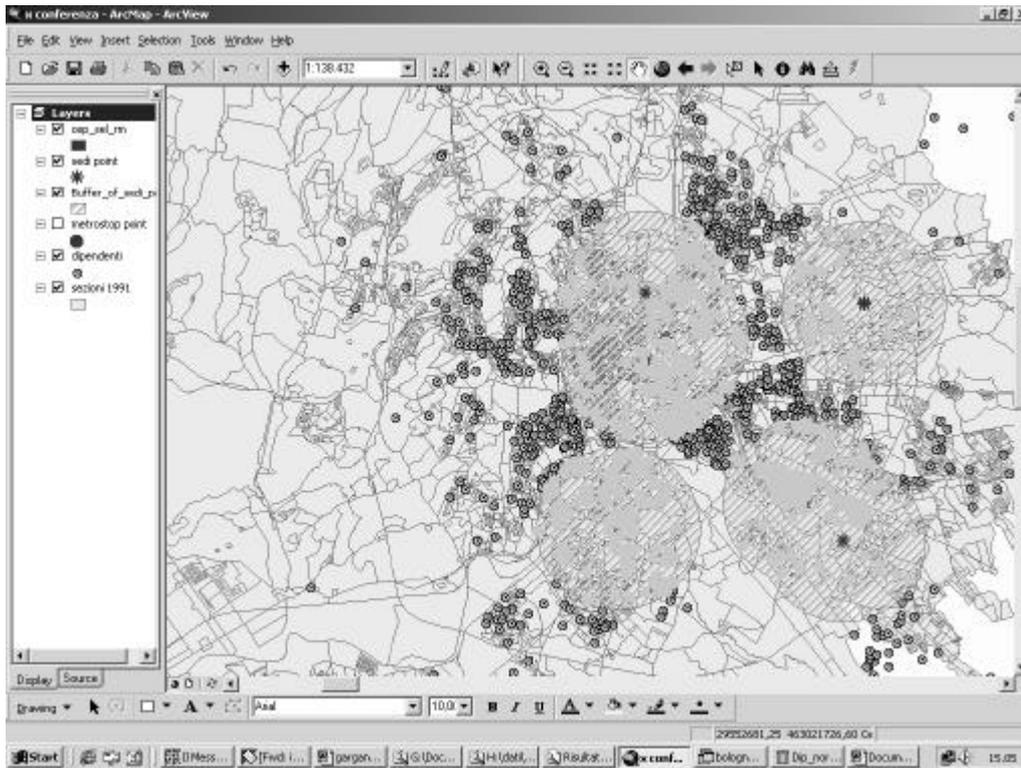
Il calcolo della distanza in linea d'aria tuttavia è un'indicazione scarsamente significativa in un contesto reale, come veniva già ricordato nel documento.

Maggiore utilità riveste il calcolo dei percorsi, possibile disponendo di un dettagliato grafo stradale, o che può essere approssimato, come è stato fatto nello studio in esame, ipotizzando gli spostamenti tramite linee pubbliche su rotaia (metropolitane o ferrovie). In Figura 6 sono rappresentate le principali linee ferrotranviarie del comune di Roma, sulle quali sono stati stimati i tempi di percorrenza dei dipendenti in funzione delle distanze calcolate con il GIS.

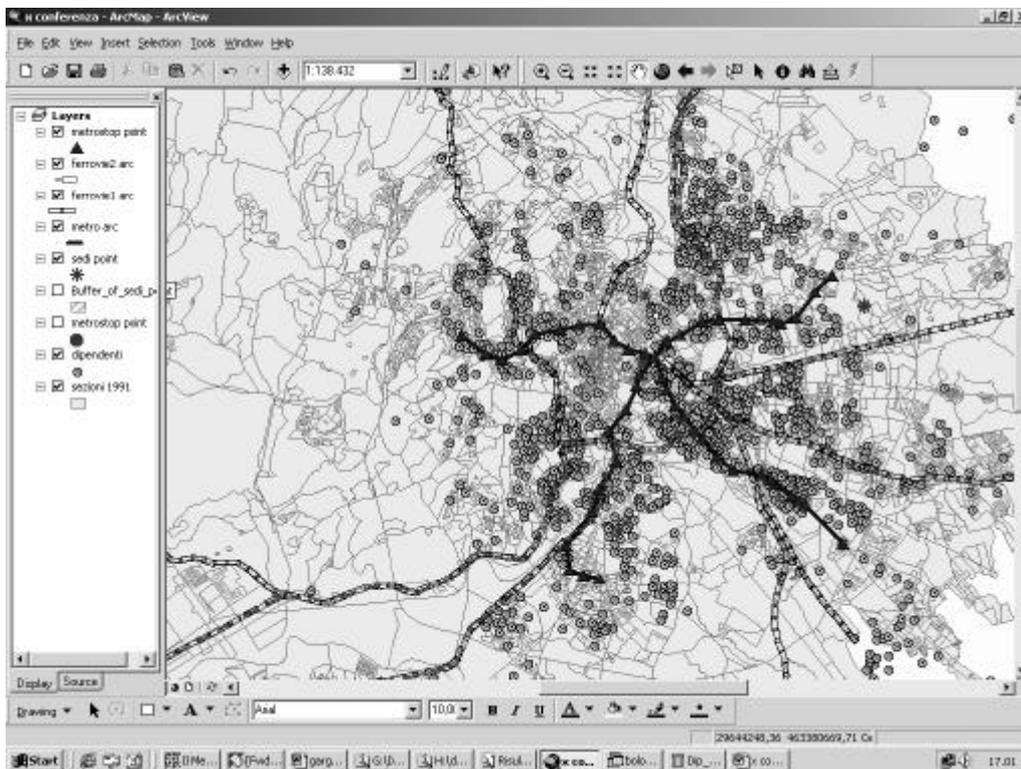
**Figura 4** *Rappresentazione dei dati*



**Figura 5** Esempio di utilizzo di funzioni GIS



**Figura 6** Rappresentazione delle principali linee di collegamento



## Riferimenti bibliografici

- Biallo G. (2002) Introduzione ai Sistemi Informativi Geografici. I *Quaderni di MondoGIS*
- Fisher, P.F., Langford, M. (1995), Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation, *Environment and Planning, A*, 27, 211-224
- Flowerdew, R., et al. (1991), Using areal interpolation methods in geographic information systems, *Papers in Regional Science*, n.70, 303-315
- Goodchild, M.F. et. Al. (1993), Aframework for the areal interpolation of socio-economic data, *Environment and Planning A*, 25, pp.383-397
- Istat (1998) Primo Atlante di Geografia Amministrativa, (Floppy disk)
- Istat (2001), Unità amministrative. Variazioni territoriali e di nome dal 1861 al 2000.
- Openshaw S. (1985), The Modifiable Areal Unit Problem, *Catmog* 38, The Invicta Press, London
- Openshaw S. (1995), *Census Users' Handbook*, Longman, London
- Orli, R. et al. (1996) Address Data Quality and Geocoding Standards. US Gov't document.
- UN/ECE (1998) *Atti della Work Session on Geographical Information Systems*, Ottawa, Canada, 5-7 October
- UNITED NATIONS (2000), Handbook on geographic information systems and digital mapping, *Studies in Methods*, Series F No.79
- Vliegen M., Van Leeuwen N. (1998), Geographical Information Systems and their Increasing Significance for the compilation of Area Based Statistics, *Seminar on New Techniques & Technologies for Statistics*, Sorrento, 4-6 Novembre