



UNIVERSITÀ DEGLI STUDI DI PALERMO
DOTTORATO IN SOCIOLOGIA, TERRITORIO E
SVILUPPO RURALE

DIPARTIMENTO SCIENZE GIURIDICHE, DELLA SOCIETÀ E DELLO SPORT

**GIS ANALYSIS NELLA
RICERCA SOCIALE
QUALI-QUANTITATIVA**

Tesi di Dottorato di:

Maria Carmela La Greca

Coordinatore Dottorato:

Prof. Fabio Massimo Lo Verde

Relatore:

Prof. Alberto Trobia

INTRODUZIONE	III
CAPITOLO PRIMO	1
SPATIALLY INTEGRATED SOCIAL SCIENCE E	1
GIS ANALYSIS	1
1.1. La SISS e la Scuola di Chicago.....	6
1.2. Ambiente, spazio e luogo	8
1.3. Analisi place-based e modellazione esplicita spaziale	9
1.4. Interdisciplinarietà: il contesto sociale della conoscenza SISS	10
1.5. Il dibattito sulla GIS analysis	11
CAPITOLO SECONDO.....	15
GEOGRAPHIC INFORMATION SYSTEMS	15
2.1. Cartografia e GIS	16
2.2. Cosa è un GIS?.....	19
2.3. Componenti di un GIS.....	20
2.4. Mappare la società: modelli di realtà.....	21
2.5. Questione di scala.....	26
2.6. Modelli cartografici.....	27
2.7. People manipulate object but live in fields. Modello vettoriale e raster.....	28
2.7.1. Formati digitali di codifica dell’informazione geografica.....	31
2.8. Georeferenziazione con Quantum GIS.....	32
2.9. Operativizzazione dei concetti nella GIS analysis: entità e attributi	38
2.10. Quando condurre una GIS analysis	43
2.11. Alcuni software GIS per l’analisi socio-spaziale.....	44
CAPITOLO TERZO.....	47
GEODATABASE.....	47
3.1. Il modello dei dati	48
3.2. Database relazionale.....	49
3.3. Vincolo di integrità.....	54
3.4. Join, proiezione e selezione.....	55
3.5. Operazioni sui dati con SQL	56
3.6. Modello georelazionale	58
3.7. Costruzione del modello concettuale e logico <i>benidiscilia</i>	59
3.8. Interrogazione dei dati geografici.....	74
3.9. Overlaying.....	79
3.10. Overlaying con ArcGIS online.....	82
CAPITOLO QUARTO.....	86
GEODATA.....	86
4.1. Trovare i geodata.....	88

Indice

4.2.	Metadata.....	94
4.3.	Accuratezza e precisione dei geodata.....	95
4.4.	Geo Big Data.....	97
4.4.1.	Geolocalizzare i tweet.....	99
4.5.	Raw OpenGeoData Now!.....	104
CAPITOLO QUINTO		111
CAQDAS E QUALITATIVE GIS		111
5.1.	CAQDAS	113
5.2.	QDA-GIS analysis.....	115
5.3.	Testi e contesti.....	116
5.4.	Qualitativeness	117
5.5.	Database ibrido-relazionale.....	119
5.6.	GIS analysis supportata da Atlas.ti.....	125
5.6.1.	Archiviare i waypoint in Google Earth.....	126
5.6.2.	Usare Google Earth su Atlas.ti.....	129
5.7.	Participatory Social Research Methods e PPGIS	133
5.7.1.	Mapuche e community mapping	134
5.7.2.	Costruzione partecipata del GIS.....	135
5.7.3.	Sincronizzazione tra <i>waypoint</i> e dati qualitativi.....	137
CAPITOLO SESTO		143
SVILUPPI FUTURI DELLA GIS ANALYSIS		143
6.1.	Geographic Information Systems e Social Network Analysis: Spatial is Special.....	143
6.2.	Spatially embedded network e autocorrelazione spaziale.....	145
6.3.1.	Dengue fever: un esempio di NNI Spatial Analysis	148
6.4.	I terabyte di Tobler.....	152
6.5.	Funzione di interazione spaziale e diagrammi di Voronoi: oltre la distanza.....	159
BIBLIOGRAFIA		164

INTRODUZIONE

Il presente lavoro ha come obiettivo generale la presentazione dell'analisi GIS (*Geographic Information System*) come un potente strumento per la ricerca sociale. Di recente, grazie al ricorso a *software* sempre più raffinati e dotati di un'interfaccia *friendly*, si comincia a registrare un suo sviluppo non solo nell'ambito delle scienze esatte e, quindi, nei settori di originaria applicazione, ma anche negli studi su territorio e socialità.

I ricercatori che analizzano i fenomeni di sociologia urbana e rurale si sono sempre posti interrogativi riguardanti l'esistenza o meno di metodologie specifiche per la rilevazione ed elaborazione dei dati territoriali. La GIS *analysis* contribuisce all'arricchimento del patrimonio di strumenti utilizzati dai ricercatori sociali [McMaster e Harvey, 2010] e, per certi versi, al rinnovamento della sociologia stessa poiché individua e delimita il contesto di analisi socio-spaziale, permette di costruire indicatori adatti a misurare i fenomeni *place-based*, tiene conto della complessità e varietà delle fonti e delle tecniche di rilevazione dei dati, della necessità di collegare i risultati ottenuti a quelli che sono i caratteri e le dimensioni del territorio in esame.

Tuttavia, perché possano sfruttare le potenzialità della GIS *analysis*, occorre che i ricercatori urbani e rurali si interrogino sullo stato della loro disciplina e, nello specifico, sulla definizione dei concetti di *spazio*, *territorio* e *luogo* [Janelle e Hodge, 2000], in riferimento sia alla città sia alla dimensione rurale. La sociologia urbana e rurale, in Italia, ha conosciuto un forte ma problematico ripensamento, risalente agli anni Novanta, nel corso del quale essa è stata definita come «la punta avanzata delle analisi sociali, sollecitate peraltro dall'impegno politico per lo sviluppo socioeconomico del Paese» [Guidicini e Sgroi, 1997]. I limitati risultati operativi, ottenuti da quel processo di ricostruzione, essenzialmente fallito, la riportano, oggi, al punto di partenza: i segni nello spazio sono assai confusi, le categorie di analisi non più condivise.

L'elaborazione dei contenuti del presente lavoro di tesi è stata animata dal desiderio di dare un contributo alla definizione di strategie metodologiche alternative che siano svincolate dagli stereotipi del passato e diano luogo a una sociologia del territorio sempre più interdisciplinare. La convergenza di nuovi tipi di dati, tecniche di analisi e dinamiche di cooperazione tra ricercatori appartenenti a discipline diverse, è la condizione necessaria allo sviluppo di un nuovo modo di fare ricerca sociale.

Dopo decenni di cambiamenti che hanno investito anche la tecnologia a supporto della ricerca sociale, la diffusione della *Spatial Analysis* [Anselin e Rey, 2010] pare dirigersi verso nuovi orizzonti. A tal proposito, lo scopo del primo capitolo è quello di delineare le peculiarità della *Spatially Integrated Social Science* (SISS) [Janelle e Goodchild, 2011]. Un approccio in persistente tensione tra teoria e pratica, non statico, ma posizionato in continuo mutamento sui due assi spazio-tempo [Kang e Scott, 2007], multidimensionale e interdisciplinare, che trascende i confini dettati dalle agende delle discipline tradizionali e che riguarda non solo le scienze umane e sociali, ma tutte le scienze in cui il rapporto uomo-territorio si rivela centrale.

Nel corso della trattazione, viene descritto il contesto nel quale la SISS continua a svilupparsi e viene delineato un quadro previsionale utile a capire quali vie stia seguendo la sua diffusione. La SISS informa la GIS *analysis* come uno dei possibili approcci teorici attraverso i quali i ricercatori sociali possono leggere le dinamiche che si manifestano nel territorio, analizzare le connessioni tra ideologie, valori e sollecitazioni che vi si sviluppano, e le modalità e strutture tecniche attraverso le quali si esprimono. Grazie allo strumento GIS, la SISS permette di esplicitare le rappresentazioni sociali degli spazi insediativi [Darra, 2009; Zlatev, 2007; Bailey e Gatrell, 1996], di conoscere il fluttuante mondo di combinazioni in cui il soggetto ritaglia i suoi meccanismi di appartenenza al territorio; non quale semplice accumulo di oggetti, eventi, cose e parole, ma luogo di senso.

Allo scopo di riassumere l'origine e gli sviluppi della SISS, si è analizzato lo stato dell'arte, aggiornandolo alla luce dei contributi raccolti a livello nazionale e

internazionale. In particolare, si è fatto esplicito riferimento a quegli autori che la considerano una prospettiva appropriata allo studio della società. L'analisi *place-based* e la *modellazione esplicita spaziale* [Fischer e Getis, 2010] sono i due perni attorno ai quali si muove la SISS, necessari alla costruzione di indicatori capaci di trasformare l'anomato spaziale in un tessuto composito definito.

L'ipotesi è che la *GIS analysis*, un vero e proprio approccio di analisi che si innesta nel macro-approccio SISS, accompagnerà perfettamente la costruzione di una teoria che si voglia dotare di potere esplicativo. Attraverso la cartografia analitica, la statistica spaziale e i GIS, lo scienziato sociale potrà integrare teoria e analisi empirica, nel tentativo di ricostruire un *sensu* sempre più complesso degli *scenari sociali* [Mela, 1992] situati spazialmente. Egli non potrà accontentarsi di proporre spiegazioni descrittive, indicando i fattori che rendono probabile o meno l'occorrenza di un fenomeno osservabile. Piuttosto, dovrà elaborare ipotesi causali verosimili sul perché un fenomeno accada, perché avvenga una variazione concomitante tra stati ed eventi nello spazio e nel tempo [Tita e Cohen, 2004]. E, nel corso di tali analisi ed elaborazioni, non potrà che fare i conti con l'incompletezza e la parzialità delle teorie. Queste ultime non potranno rifuggere dall'importanza del mondo sociale ma, allo stesso tempo, non avranno la pretesa di rappresentarlo così come è.

Da un lato, il *Quantitative GIS Approach* [Crampton, 2009] sarà in grado di ridurre la complessità del reale, semplificandola, e di non spingere il ricercatore nella trappola della razionalità lineare che offrirebbe solo una debole corrispondenza numerica delle cause e degli effetti. Dall'altro lato, il *Qualitative GIS Approach* [Al-Kodmany, 2002; Shiffer, 2002; Weiner e Harris, 2003; Fielding, 2010] avrà il vantaggio della chiarezza rappresentativa di una lente d'ingrandimento che non sacrifica la validità predittiva ed esplicativa del modello logico e concettuale GIS, poiché di esso faranno parte tutti i caratteri e le dimensioni simboliche del territorio in esame, rilevati nel corso dell'indagine qualitativa.

Nel secondo capitolo, si è data la definizione di GIS descrivendo le diverse componenti che lo costituiscono. I contributi sono stati raccolti principalmente

nell'ambito della letteratura riguardante l'ingegneria dell'ambiente e del territorio e della cartografia numerica [Dainelli *et al.*, 2010].

Le carte sono sempre state, anche per i sociologi, un utile strumento di conoscenza e controllo del territorio. Tuttavia, elaborarle è un'operazione tecnica difficile, che richiede specifiche conoscenze professionali trasversali. Non basta osservare o descrivere un territorio; occorre farlo seguendo le regole che stanno alla base dell'elaborazione dei modelli GIS logico, concettuale e fisico.

Il terzo capitolo approfondisce proprio le procedure di elaborazione del modello fisico: il geodatabase [Steinberg e Steinberg, 2006]. Riportando alcuni esempi di *modellizzazione esplicita spaziale*, sperimentata nell'ambito di indagini condotte nel corso dell'esperienza di dottorato, sono descritte le modalità di inserimento nel geodatabase di tutte le proprietà topologiche (rapporti di contiguità, vicinanza, inclusione ed esclusione territoriale, ecc.) di cui godono le entità spaziali identificate e che, in genere, sono assenti nei database tradizionali. È in questa fase di *building* [Klathen, 2007; Lee *et al.*, 2005] che il ricercatore potrà scegliere le regole che gli permetteranno di interagire con la rappresentazione cartografica, di interrogarla e modificarla, in base alle esigenze di analisi.

Dopo aver approfondito le strategie di combinazione dei diversi *dataset* e avere sottolineato quanto sia indispensabile affrontare il processo di operativizzazione tenendo conto dei vincoli che dipendono dalla natura del modello concettuale e logico elaborato, il quarto capitolo passa in rassegna le fonti di rilevazione dei dati spaziali concentrandosi sulla raccolta dei geodata secondari [Martin *et al.*, 2012].

Non ci si rivolge al ricercatore sociale GIS come se questi si trovasse nel bel mezzo del deserto conoscitivo, costretto a dover affrontare questioni a cui nessun altro ha mai pensato, a rilevare dati mai disponibili e che nessun altro ha mai prodotto o rilevato. Piuttosto, coerentemente con quanto avviene nella realtà, si ritiene che il percorso di ricerca GIS non debba essere ingessato in un modello tipico e che la cumulatività delle informazioni disponibili favorisca la loro fruizione creativa.

La facilità con cui i geodata vengono prodotti e accumulati con elevata frequenza, e il modo in cui vengono resi accessibili con cadenza regolare presso una moltitudine di archivi *online* e *offline*, li connota come un patrimonio informativo insostituibile per lo studio delle stesse comunità che li producono. La loro analisi costituisce un'opportunità sia per la verifica e la revisione di asserti, prodotti da altri osservatori GIS, sia per la risposta a domande non individuate e/o a questioni non ancora affrontate [Macintyre *et al.*, 2002; Openshaw, 1984]. L'analisi dei geodata secondari ha il potenziale di rendere fallibile la GIS analysis primaria altrui o, meglio ancora, la propria. È altamente desiderabile, infatti, per il bene della scienza sociale, che il ricercatore metta in discussione i suoi stessi tentativi congetturali e le sue falsificazioni, pubblicizzando i geodata originali e indicando ad altri quali sono gli strumenti per aggregarli o modificarli. Aggiungere, a un *geodataset* esistente, i diversi dati che il ricercatore ha raccolto per la sua analisi, provenienti da diverse fonti, può essere complicato. Ma permette di incorporare informazioni eterogenee, normalmente difficili da integrare tra loro, in un ambiente di analisi unico. Occorre scegliere i geodata in maniera pertinente mantenendo saldo il principio dell'indissolubilità tra teoria e ricerca, tra metodo e tecniche. Non è certo che tale modo di procedere sia una garanzia di validità ma, scegliere di farlo, può rappresentare un ingente risparmio e spesso l'unica possibilità di studiare il problema in maniera comparata, migliorando l'attendibilità dei geodata stessi. È stato evidenziato, inoltre, come una ricerca sociale sempre più *cross-national*, e che manifesta chiare difficoltà legate al reperimento dei finanziamenti, necessiti d'investire su cambiamenti sostanziali a livello organizzativo e di formare imprese collettive interessate ai dati come bene condiviso [Biolcati-Rinaldi e Vezzoni, 2013].

Il capitolo si conclude approfondendo le strategie di verifica dell'attendibilità dei geodata poiché - come detto - le tecniche di ricerca sociale non migliorano la qualità dei dati stessi.

Il valore aggiunto della tecnologia GIS e delle tecniche di analisi spaziale aumenta in maniera direttamente proporzionale alla creatività del ricercatore che intende utilizzarle per i propri studi. Esistono numerose tecniche che possono essere

integrate ai GIS. Nel corso della trattazione, si spiegherà come poterne impiegare alcune allo scopo di misurare l'influenza reciproca tra spazio e azioni. Si distinguerà, infatti, tra percorsi di ricerca diversi attraverso i quali misurare le associazioni spazio-temporali per testare le ipotesi e produrre vantaggi e benefici per la ricerca sociale.

In particolare, nel quinto capitolo, si è discusso dei primi tentativi di integrazione tra dati qualitativi e GIS, effettuati da quei ricercatori di diverse discipline (antropologi, psicologi e geografi umani) che cominciarono a riflettere sulle potenzialità di questo connubio, contribuendo alla messa a punto di quello che è poi diventato il *Qualitative GIS Approach*. Un approccio che è molto più che una semplice integrazione tra *software* QDA e GIS [Elwood, 2006, 323-341; Crang, 2005, 225-233]. I ricercatori che se ne avvalgono assumono che l'interpretazione dei dati utili all'analisi venga prodotta e negoziata seguendo un processo non lineare, ma iterativo e circolare [Seidel, 1998, 2], soggetto a continue modificazioni.

Se, dunque, nella prima parte della tesi, si discute di come l'analista GIS possa percorrere, in successione, le fasi standardizzate del processo di definizione del modello concettuale e logico fino all'elaborazione del geodatabase fisico, qui si approfondiscono le diverse modalità attraverso le quali egli possa arricchire il geodatabase con dati eterogenei, opportunamente codificati. Seguendo il *Qualitative GIS Approach*, il ricercatore potrà più volte tornare sui suoi passi, dalla fase di raccolta dati, e di una concomitante interpretazione degli stessi, fino al raggiungimento della saturazione teorica attraverso la produzione di teorie concettualmente dense [Cicognani, 2002]. Il processo di codifica *preliminare*, che caratterizza la *GIS analysis*, verrà continuamente messo in discussione dalla codifica *a posteriori*, tipica della ricerca qualitativa. Il ricercatore, da un lato, avrà la possibilità di riprodurre preliminarmente la realtà sociale sistematizzando, in un ambiente GIS, i dati che ad essa si riferiscono; dall'altro, grazie alle comparazioni cicliche che caratterizzano il processo di analisi qualitativo, potrà occuparsi di riassembleare tali dati e di produrre il modello teorico-interpretativo. La stessa definizione di approccio qualitativo ai dati, del resto, è di tipo aperto perché, pur

prevedendo la necessità di interpretarli, non esclude che, anche nel processo di codifica, si faccia riferimento al dato quantitativo [Lewins e Silver, 2004]. In egual modo, le regole di un'analisi GIS di tipo *standard*, si sposano bene con quelle previste dall'analisi qualitativa computer-assistita, dato che anche il ricercatore che si avvale dei QDA dovrà formulare delle ipotesi iniziali, seppur soggette a revisioni continue.

Il *Qualitative GIS Approach* lascia emergere la riduttività di una visione dicotomica che definisce i modelli qualitativo e quantitativo come totalmente separati e opposti [Kruglanski e Jost, 2000, 45-67]. Collocandosi nel cosiddetto *continuum* qualità-quantità [Denzin, 1978], a difesa del pluralismo metodologico [Mannetti, 1998], facilita l'applicazione di strategie che spingono il ricercatore verso continui aggiustamenti, nei diversi momenti della ricerca. Supporta le sue capacità creative e interpretative, sin dall'inizio del processo di analisi, in cui avviene la sistematica organizzazione dei dati raccolti; facilita l'esplicitazione di procedure rigorose, indispensabili sia all'ispezionabilità della base empirica [Strati, 1997; Mazzara, 2002, 21-41], che all'utilizzo di entrambi i tipi di *software*. Ma il vantaggio più rilevante del *Qualitative GIS Approach* sta nella possibilità di condividere e confrontare il lavoro tra ricercatori diversi che usano *software* diversi, nella fattispecie QDA e GIS, purché seguano le stesse procedure in maniera sistematica. Ciò, a conforto del rigore metodologico che, nelle scienze sociali, si manifesta anche grazie alla collaborazione di gruppo e alla multidisciplinarietà. L'integrazione QDA-GIS, in particolare, costringe sia all'accordo tra diversi ricercatori sui significati da assegnare alle categorie, sia al confronto tra diversi sistemi di codifica.

Nel corso del sesto capitolo, dedicato agli sviluppi futuri della GIS *analysis* nella ricerca sociale, si è discusso di alcune delle tecniche statistiche di modellizzazione dei *network* di vicinato. Del resto, gli approcci utili ad investigare lo spazio fisico e relazionale dimostrano una forte sovrapposizione concettuale [Park *et al.*, 1925; Snow, 1854; Abbott, 1997] che può sintetizzarsi con l'assioma fondamentale della geografia o *Prima Legge di Tobler* [Tobler, 1970]: le entità spaziali sono molto simili tra loro ma quelle vicine geograficamente lo

sono ancora di più. A misurare tale somiglianza, l'autocorrelazione spaziale [Moran, 1950], il più importante concetto nell'ambito delle statistiche spaziali. La maggior parte dei fenomeni spaziali è spazialmente autocorrelata; nelle aree metropolitane, gli indicatori socio-demografici, come la densità di popolazione, il livello di alfabetizzazione, la condizione professionale, l'immigrazione, i servizi collettivi, sono distribuiti in modo non casuale nello spazio. Inoltre, gradi diversi di autocorrelazione spaziale positiva si ritrovano a tutte le scale di osservazione (regionale, nazionale, globale).

Tuttavia, nonostante le chiare ed evidenti sovrapposizioni concettuali, l'integrazione formale tra *social network* e strategie analitiche spaziali resta relativamente poco affrontata in letteratura. Perciò, si è ritenuto necessario puntare l'attenzione sulle modalità di identificazione delle direzioni future della combinazione tra SNA e analisi spaziale [Adams, Faust e Lovasi, 2012; Marsden, 2011; Hipp, 2012; Barthelemy, 2011].

Il capitolo si conclude con l'esame di un ulteriore tipo di integrazione: GIS-ABM. Si ritiene che la modellazione basata su agenti e gli strumenti di analisi offerti dai GIS possano orientare il ricercatore nella comprensione delle complesse dinamiche d'interazione sociale. Del resto, anche i modelli ad agente, così come i modelli GIS, sono costituiti da *entità*, caratterizzate da *proprietà* o *attributi*. E, dal momento che è possibile manipolare a piacere alcuni parametri del modello, agire su quelli geo-spaziali potrebbe facilitare l'osservazione delle conseguenze di tali scelte. Dopo aver esplorato i limiti riguardanti la scelta dei nodi localizzati sulla base *landscape*, ci si è concentrati su come la distribuzione di questi sia influenzata dagli schemi d'uso del territorio da parte degli agenti che si muovono e scelgono in base a criteri di prossimità e preferenza. Gli agenti sono in grado di interagire tra loro e con l'ambiente circostante, dando origine a fenomeni sociali complessi e modificando il loro comportamento futuro [Boero e Squazzoni, 2005].

La trattazione, nel suo complesso, non rappresenta affatto un punto di arrivo. Piuttosto, un'intensa incursione; un punto di partenza per ulteriori esplorazioni delle dinamiche del sottosistema delle comunicazioni sociali che si

manifestano nel territorio. È anche un invito al rigore terminologico, perché si eviti di usare un'espressione del tutto vaga come *spazio*. Altrimenti, ogni punto geolocalizzato diventerà equivalente ed estraneo, sacrificato sulla croce degli assi cartesiani [de Certeau, 2001].

CAPITOLO PRIMO

SPATIALLY INTEGRATED SOCIAL SCIENCE E GIS ANALYSIS

L'analisi spaziale, da tempo, ha trovato largo eco in seno all'interesse di ricercatori di natura diversa: sociologi, antropologi, psicologi, economisti, epidemiologi, geografi. I sociologi ne hanno sfruttato le potenzialità analizzando la dimensione, densità e omogeneità dei processi sociali situati spazialmente [Martinotti, 1993; 1999; Mela *et al.*, 2000; Nuvolati, 2002; 2007; Park *et al.*, 1925]; gli psicologi hanno concentrato l'attenzione sulla variabile comportamentistica e sul sistema delle interazioni sociali che si manifestano nello spazio [Logan e Collver, 1983]; gli antropologi hanno introdotto come elementi chiave dell'analisi le variabili *tradizione* e *cultura*, analizzando la stratificazione storica, i caratteri dell'esperienza che fanno della realtà sociale una unità con una specifica struttura demografica e spaziale, un insieme di usanze e di idee, di termini di giudizio e comportamenti comuni [Banfield, 1976]; gli economisti hanno analizzato i modelli di produzione presenti nell'area, la struttura occupazionale e i consumi prevalenti¹ [Florida, 2005; Lynd e Lynd, 1970]; i geografi hanno approfondito la relazione uomo-ambiente attraverso lo studio della distribuzione dei fenomeni sul territorio, ponendo l'accento su variabili nodali quali *prossimità fisica* e *distanza sociale* [Wirth, 1945]; gli epidemiologi hanno analizzato le caratteristiche naturali e antropiche dell'ambiente fisico, al fine di rappresentare lo scenario in cui ricercare i fattori di rischio per la salute collettiva e individuale [Snow, 1854].

Nel corso di quasi tutto il Novecento, i confini fluttuanti e incerti dello spazio consumato, abitato, visitato hanno motivato le ricerche sociologiche empiriche che hanno finito per raccontarlo quasi sempre come ferito, vuoto, complesso, contraddittorio, delineato da aree di conflittualità, disomogeneità. La

¹ Spesso, nelle ricerche di sociologia urbana e rurale, accade che il taglio antropologico lasci lo spazio a quello dedicato allo studio sulla struttura occupazionale, i modelli di produzione presenti nell'area di interesse, i consumi prevalenti, i modi di impiego del tempo libero.

storia della sociologia urbana e rurale ha inizio proprio quando lo spazio, inteso come variabile indispensabile alla lettura dei fenomeni sociali, venne messo in crisi in quanto entità teoricamente significativa, espressione e forma dei processi sempre mutevoli di accumulazione capitalista [Marx, 1859], o base dell'associazione umana, dell'identità sociale e della dominazione politica [Weber, 1961], o della divisione del lavoro [Durkheim, 1893]. Restituire allo spazio la valenza esplicativa persa è stato, da sempre, il compito di una disciplina che non ha mai offerto casi studio generalizzabili.

La difficoltà della sociologia urbana e rurale è determinata dall'inesistenza di un modello interpretativo sistematico che spesso spinge il ricercatore a rappresentare l'urbano e il rurale come una sequenza di impressioni, o un accostamento di dati.

Al di là delle risposte, la domanda che tormenta ancora i sociologi urbani e rurali è la stessa, dall'inizio del secolo scorso: «Come gli spazi si fanno luoghi?».

I realisti naturali sono ricorsi al *genius loci* [Norberg-Schulz, 1979], quello spirito del luogo che si manifesta in maniera costante nel tempo e che vincola e unisce emotivamente i soggetti che vi vivono; i costruttivisti, al *carattere del luogo* [Massey, 1993], che mescola continuamente i suoi significati in funzione delle pratiche sociali, culturali ed economiche che influenzano le azioni dei soggetti e dei gruppi [Nuvolati, 2011].

Qui, non si celebrerà l'onnipotenza dell'approccio spazialista [Mela, 2006] che risalta le capacità predittive delle componenti fisiche dei luoghi. Piuttosto, si ricorderà come queste ultime siano spesso indipendenti dagli aspetti spaziali, i quali, a loro volta, possono concorrere alla determinazione degli stessi fenomeni indipendentemente dall'unicità dei luoghi [Nuvolati, 2007]. «*The sense of place* non è qualcosa che noi dobbiamo sentire del posto, ma qualcosa che il posto deve "consentire" di darci (...). I luoghi sono presenze quindi hanno l'istantaneità e l'imprevedibilità delle presenze» [La Cecla, 2000, 150-151].

La natura di tali forme e strati che mutano sembianze e, allo stesso tempo, resistono alle trasformazioni può essere spiegata attingendo agli strumenti offerti dal terreno privilegiato di impegno analitico sul quale la SISS (*Spatially Integrated Social Science*) affonda le sue radici. Essa giustifica un sistema di

regole che stabilisce ciò che i sociologi, interessati alla spiegazione dei fenomeni situati spazialmente, debbano fare piuttosto che ciò che debbano dire di fare. Per la SISS, la questione teorica diventa, di fatto, metodologica. Le definizioni e le specificazioni di significato stipulate nell'ambito di tale approccio fanno parte di un contratto in cui non viene presentato il resoconto delle scoperte ma delle modalità tecniche per poterle compiere.

I ricercatori che analizzano i fenomeni di sociologia urbana e rurale si sono sempre posti interrogativi riguardanti l'esistenza o meno di metodologie specifiche per la rilevazione ed elaborazione dei dati territoriali. Un'adeguata metodologia legata ai fenomeni del territorio non può, ad esempio, non tenere conto della specificità del contesto, che deve essere adeguatamente individuato e delimitato, della necessità di costruire indicatori adatti a misurare i fenomeni *place-based*, della complessità e varietà delle fonti e delle tecniche di rilevazione dei dati, della necessità di collegare i risultati ottenuti a quelli che sono i caratteri e le dimensioni del territorio in esame.

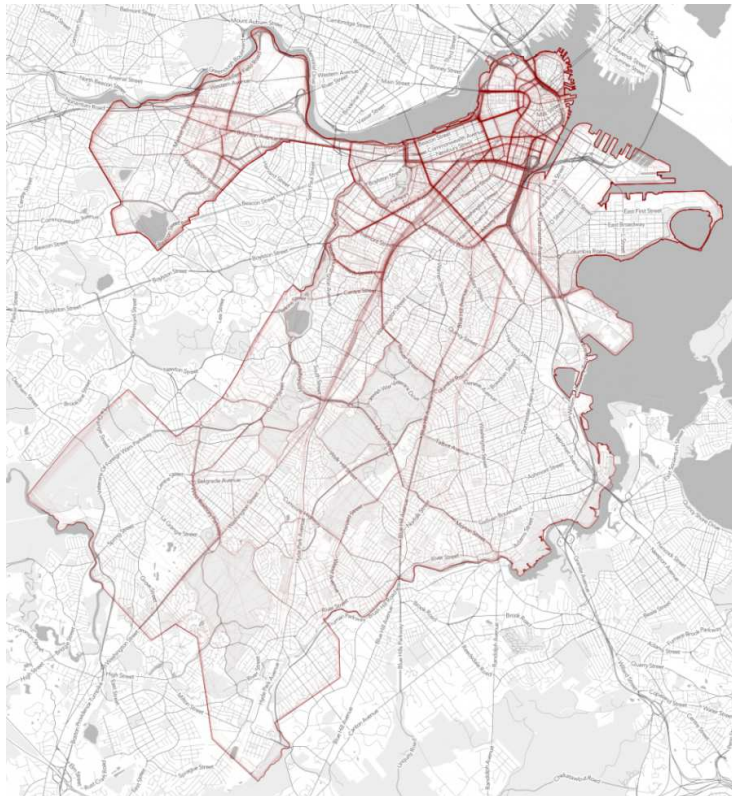
Sin dalla nascita della SISS, i suoi principali sostenitori hanno giovato dell'apporto di altri saperi che offrono metafore, concetti, tecniche necessari a spiegare il significato che lo spazio acquisisce in relazione con l'esperienza individuale e con il suo investimento simbolico da parte dei gruppi e delle comunità che vi vivono. Ogni oggetto d'uso localizzabile nello spazio denota una funzione, un'attività e l'insieme delle valenze ideologiche e psicologiche che si collegano alla funzione stessa. Ogni percorso, nodo, margine, zona può essere interpretato come messaggio funzionale allo spostamento del soggetto all'interno dello spazio [Agustoni, 1997, 90].

Alla SISS si riconducono quelle tecniche di analisi che permettono di studiare i gruppi sociali locali e globali, le similitudini e le diversità territoriali, le clusterizzazioni delle entità spaziali e delle loro proprietà. L'applicazione di tali strategie di analisi passa dalla consapevolezza, da parte del ricercatore, che tali entità non debbano georiferirsi² esclusivamente entro contesti delimitati da confini

² Sul concetto di georeferenziazione, si rimanda al Capitolo Secondo.

politici o fisici, ma anche dalle percezioni di chi vi abita, che vi attribuiscono un senso.

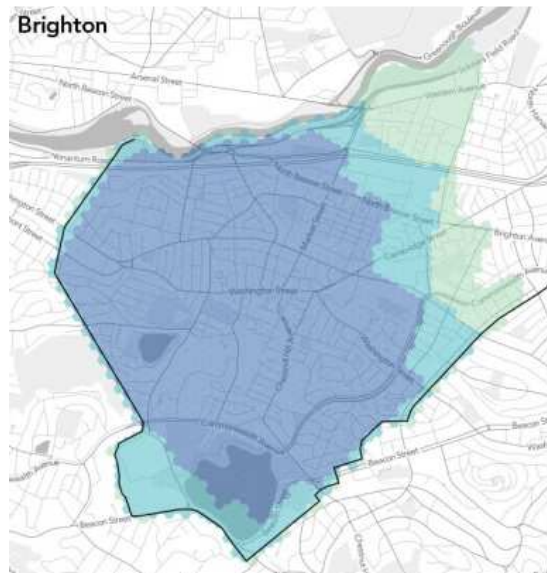
Nella celebre ricerca *Bostonography*, Woodruff e Wallace [2012], dell'Università di Wisconsin-Madison, hanno costruito una mappa interattiva della città di Boston, chiedendo ai residenti di indicare i confini percepiti del loro quartiere. Ogni qualvolta tracciano lo stesso confine, questo viene rappresentato con una linea più marcata. L'elaborazione cartografica è piuttosto imprecisa ma, certamente, è una trasposizione fedele di come i soggetti, coinvolti nella ricerca, percepiscano i confini dei quartieri in cui vivono. Se si osserva bene, si noterà come la mappa presenti diverse intersezioni tra i poligoni disegnati.



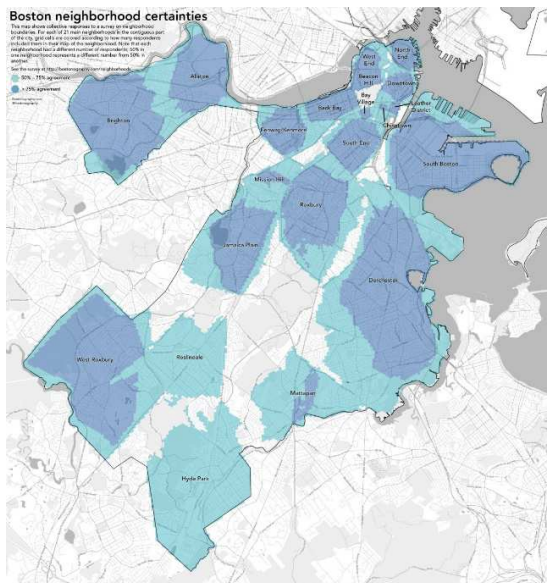
(Figura 1.1. Web-based GIS map dei confini dei quartieri di Boston, tracciati dai residenti i quartieri stessi. Fonte: <http://bostonography.com>).

Nella mappa rappresentata in Figura 1.2., si può notare come, per una certa percentuale di residenti coinvolti nella ricerca (circa il 25%), il confine estremo

orientale di Brighton valichi quella che, in base alla cartografia amministrativa, è l'area del quartiere limitrofo di Allstone.



(Figura 1.2. Confini percepiti dagli abitanti del quartiere di Brighton, Boston. Fonte: <http://bostonography.com>).



(Figura 1.3. Confini percepiti dai residenti dei quartieri di Boston. Fonte: <http://bostonography.com>)

Interessante citare, a proposito dello studio sui quartieri, le analisi di Jacobs [1958], che pure hanno ispirato i ricercatori SISS come tutti i sociologi

urbani e rurali. Secondo la studiosa, il proliferare di reti di relazioni informali è determinato anche dalle caratteristiche degli spazi pubblici. Le condizioni che generano diversità, di persone, culture, edifici, attività generano un apparente disordine urbano che viene collegato alla vitalità. La commistione tra le funzioni di produzione e residenza debbono poter assicurare la circolazione di persone che popolino le strade a ore diverse e che abbiano modo di utilizzare in comune i beni e i servizi urbani. Il limite fisico entro il quale questi bisogni vengono soddisfatti, influenzando positivamente o negativamente lo sviluppo della diversità e della vitalità, spesso coincide con i confini percepiti dai residenti il quartiere. È anche in base a tali esperienze di studio e ricerca che i sociologi SISS concludono che affermare che lo spazio è socialmente formato sia alquanto riduttivo poiché, spesso, questo assume una forma che sfugge al controllo empirico e genera esso stesso socialità.

La SISS non studia lo spazio come un oggetto a sé o una variabile indipendente. Si tratta di un approccio tutt'altro che statico. Piuttosto, è posizionato in continua tensione tra teoria e pratica. È multidimensionale e multidisciplinare poiché trascende i confini dettati dalle agende delle discipline tradizionali e i suoi strumenti sono utili non solo agli studiosi di scienze umane e sociali ma di tutte le scienze in cui il rapporto uomo-territorio è centrale. La cartografia numerica, la statistica spaziale e la GIS *analysis* sono tra i più significativi strumenti SISS in grado di supportare l'analista nell'esecuzione di operazioni analitiche utili ad esplorare sistematicamente i processi comunicativi che si manifestano sul territorio [Mela, 1992] e a confrontare spazio epistemologico ed empirico.

1.1.La SISS e la Scuola di Chicago

L'origine della SISS viene fatta risalire intorno agli anni Novanta e al lavoro dei ricercatori che ne hanno determinato la diffusione, fino ai tempi più recenti [Longley *at al.*, 2010; Janelle e Goodchild, 2011; Steinberg e Steinberg, 2006; Nyerges, Couclelis e McMaster, 2011]. Tuttavia, prima di allora, alcune riflessioni sull'importanza della variabile *spazio*, nello studio dei fenomeni sociali

spatially-embedded, erano già state formulate da diversi gruppi di ricerca, compreso quello della Scuola di Chicago. Le tecniche di studio di cui si avvalsero i ricercatori progressisti della Scuola, quali l'analisi della distribuzione dei fenomeni sociali sul territorio o delle trasformazioni urbane, sono molto affini a quelle proposte dalla SISS.

Tuttavia, in tanti, tra ricercatori SISS e non solo, hanno riconosciuto che l'utilità del concetto di *area naturale*, tanto caro alla Scuola, tendesse a decrescere nel tentativo di formalizzarlo sulla base di indicatori empirici quantificabili [Zorbaugh, 1995; Hatt, 1946].

Per la scuola ecologica, gli individui occupano il loro spazio all'interno delle cosiddette *aree naturali*, caratterizzate da un senso d'urbanità quasi primitivo [Wirth, 1945]. I gruppi si distinguono per la loro precisa collocazione spaziale, in base alle risorse disponibili e all'esito della competizione per il loro ottenimento. La città si sviluppa, perciò, in seguito a dinamiche conflittuali tra gruppi sociali che lottano per posizionarsi su uno spazio in grado di soddisfare meglio i propri bisogni. Il cosiddetto equilibrio biotico [Gettys, 1961, 99] si traduce in una relazione spaziale relativamente stabile che si ritrova in ogni popolazione.

Non vi è dubbio che l'interesse per lo studio delle aree naturali abbia avuto una certa rilevanza nello sviluppo dei principi della SISS: rappresentò il primo passo verso lo studio della struttura interna della città, dei meccanismi sociali sottostanti agli aspetti visibili dell'organizzazione urbana, dei caratteri di ordine e regolarità che si manifestano nello spazio. Tuttavia, il principale errore dell'approccio ecologico classico, superato dalla SISS, sta proprio nell'aver cercato di individuare dei principi interni alla distribuzione territoriale, in grado di spiegare, mediante l'analisi della struttura spaziale della città, la struttura sociale sottostante [Martinotti, 2011, 114].

Insomma, se il rischio della scuola ecologica è quello di restringere troppo le problematiche a una pura e semplice scomposizione del territorio in *aree naturali*, la portata euristica della SISS dipende, invece, soprattutto da quanto l'unità di analisi scelta rappresenti un'area territoriale significativa per il fenomeno oggetto di studio, in funzione di caratteri più complessi (senso di appartenenza, desiderio

di inclusione/partecipazione, ecc.), non sempre misurabili in termini di omogeneità. Le tecniche SISS, quali la GIS *analysis*, facilitano l'immersione nella densa ed eterogenea scena urbana e rurale, costituita da una cangiante morfologia di colori sociali dello spazio e influenzata dai gruppi che agiscono, in un contesto territoriale, in relazione ad altri gruppi.

Tali tecniche offrono al ricercatore la possibilità di ricostruire tale scena rintracciando nello spazio oggetti corporei più che astratti, trasferibili nel geodatabase sotto forma di dati qualitativi e quantitativi, e di tenere conto sia della struttura sociale che delle relazioni di questa con i fattori geo-spaziali.

1.2. Ambiente, spazio e luogo

Gli studiosi SISS si sono interrogati, e si interrogano tuttora, sulla definizione dei concetti *spazio*, *territorio* e *luogo*, in riferimento sia alla città che al rurale, cercando di dar ordine a un filone di studi su cui ormai convergono forti interessi, sia istituzionali che di mercato.

Usano il termine *ambiente* per indicare un contesto fisico in accezione topografica. È un contesto privo di componenti relazionali e interattive e può essere analizzato a partire da una prospettiva oggettiva. Indica l'insieme delle condizioni all'interno delle quali qualcosa esiste o accade [Wapner e Demick, 2002]. L'ambiente gode di una dimensione simbolica poiché evoca e rivela le contraddizioni dello spazio sociale. Tra quest'ultimo e lo spazio geografico-ambientale c'è una quasi-identità. Nella coesistenza di tali dimensioni, i gruppi sono in grado di percepire l'illusione dell'unicità e di impegnarsi a trasformare il proprio spazio in *luogo*, riflesso dell'identità [Mela, 1992]. I *luoghi* possono essere studiati concentrando l'analisi sul contributo cognitivo, affettivo e comportamentale delle persone che li abitano. Sono spazi geografici antropomorfizzati, ambienti le cui caratteristiche dipendono da ciò che le persone hanno immaginato, costruito, modificato al loro interno. Insomma, sono il risultato di un processo di lunga durata che intreccia elementi spaziali e simbolici, connessi alla storia e alla cultura dei gruppi sociali che vi si sono insediati.

Mondi vitali di persone, luoghi di esperienze, caratterizzati da confini percepibili, sui quali gli individui hanno controllo, o aspirano ad averlo,

difendendoli da minacce e tutelando l'identità che in essi si è affermata³. Gli individui *immaginano* i propri luoghi, li percepiscono e li rappresentano creando un legame affettivo. L'identità di luogo, insomma, è una sub-struttura dell'identità del Sé composta da ricordi, credenze, attitudini, valori, preferenze. Rappresenta simboli, si colloca nella matrice storico-culturale delle relazioni tra gruppi, è investita di significati sociali.

Molti altri concetti, oltre quelli citati, sono stati oggetto di ridefinizione da parte degli studiosi della sociologia integrata spazialmente.

1.3. Analisi place-based e modellazione esplicita spaziale

La SISS si caratterizza per il trattamento dei *Geographic Cross-Sectional Data*, ovvero dati di natura diversa, collegabili grazie ad un minimo comune denominatore, la loro localizzazione geografica. Lo spazio georeferenziato rappresenta la fonte privilegiata per la comprensione e risoluzione non solo di problemi di spiegazione ma anche di interpretazione e comprensione. Le località possiedono anche una propria dimensione culturale: rappresentano i teatri entro cui si svolgono le interazioni e che gli individui usano per garantire un significato agli atti comunicativi [Giddens, 1981]. Le localizzazioni non sono assimilabili a semplici punti spaziali. «Possono essere stanze di una casa, angoli di strada, reparti di fabbrica, città grandi e piccole, fino alle aree, demarcate territorialmente, occupate dagli stati nazionali» [*ibidem*, 117-118]. Il dato localizzato è la componente volta ad esplicitare il fluttuante mondo di combinazioni in cui il soggetto ritaglia i suoi meccanismi di appartenenza al territorio, quale non semplice accumulo di eventi, cose e parole ma luogo di senso.

Quando il ricercatore SISS interpreta le forme georeferenziate, anche grazie alla componente ludica del *software*, non si aspetta di trovare soluzioni ma di vedersene suggerire. Quando vengono analizzati dati *cross-sectional*, è comune che l'*output* grafico mostri la residualità di un modello che non rappresenta perfettamente la realtà. La speranza è che possa suggerire fattori che possano

³ I *terrain vague* sono quelli su cui gli individui non possono esercitare alcun tipo di controllo e nei confronti dei quali non possono sviluppare legami psicologici significativi.

spiegare meglio la variazione residuale e migliorare così il modello. L'analisi *place-based*, insomma, si basa su un criterio: ogni modello o teoria nelle scienze sociali fallirà nel riportare i fenomeni che intende descrivere. Tuttavia, le possibilità del fallimento si riducono se si considera lo spazio come il necessario *link* che mette insieme dati eterogenei e provenienti da fonti diverse.

L'attributo spaziale delle entità sociali, spesso implicito nella cornice teorica adottata dai sociologi, può essere espresso attraverso modelli formalizzati che legano teorie e processi situati geograficamente. La modellazione esplicita spaziale prevede l'inserimento nel database di tutte le proprietà topologiche e di prossimità, dei rapporti di contiguità, vicinanza, inclusione ed esclusione territoriale che in genere sono assenti nei database tradizionali. Incorporare i modelli spaziali nell'analisi sociologica genera nuove opportunità di comprensione dei processi sociali stessi.

1.4. Interdisciplinarietà: il contesto sociale della conoscenza SISS

Spazio ed entità spaziali non sono oggetto di studio della sola scienza geografica. Nel 1990, Giddens scriveva: «Normalmente, gli scienziati sociali si sono accontentati di lasciar fare agli storici gli specialisti del tempo e ai geografi gli specialisti dello spazio, mentre loro mantenevano una propria identità disciplinare distinta che o aveva un interesse esclusivo per il vincolo strutturale o era focalizzata, concettualmente sulla società. Gli storici e i geografi, per parte loro, erano abbastanza conniventi verso questa dissezione disciplinare della scienza sociale. (...). Così, la storia può essere vista come una cosa che ha a che fare con successioni di eventi disposti cronologicamente nel tempo o, magari, ancora più ambigualmente, col passato. Quanto alla geografia, molti geografi amano dire che essa trova il suo carattere distintivo nello studio delle forme spaziali. Ma, se (...) le relazioni spazio-temporali non possono essere tirate fuori dall'analisi sociale senza minarne l'intera impresa, simili divisioni disciplinari impediscono attivamente di affrontare questioni di teoria sociale che siano significanti per le scienze sociali nel loro complesso [Giddens, 1990, 277-278].

Di fronte all'emergere di problematiche e orientamenti teorici e tecnici nuovi, appare evidente che la sociologia urbana e rurale debba cambiare passo e far propria un'ottica multidisciplinare molto più accentuata rispetto a quella adottata sinora. Diversamente, accanirsi sulla divisione del lavoro scientifico rischia di funzionalizzare il pensiero e rendere corrosiva qualsiasi analisi sociale, al di là delle intenzioni del ricercatore. Di fronte alla mancanza di aspettativa nei riguardi della scienza e soprattutto dell'integrità degli scienziati sociali, la SISS è la prova che la coesione tra cosmi intellettuali diversi è possibile. Anzi, essa fonda la sua esistenza proprio su tale sua potenzialità.

La SISS permette ai ricercatori sociali di spingersi oltre le applicazioni di tecniche tipiche della sociologia, fino alla più profonda comprensione dei processi sociali. Riconosciuta la necessità di agire congiuntamente, occorre che gli studiosi appartenenti ai diversi settori di ricerca stilino un programma in modo da fissare solide basi per il confronto tra tecnici e scienziati sociali. Ma, se si riscontra un rigetto e l'impossibilità a costituire équipe simili, allora questo potrebbe essere un segno significativo che va a definire la diagnosi, già anticipata da Merton [1949] sullo stato raggiunto dalla scienza postmoderna.

1.5. Il dibattito sulla GIS analysis

Il dibattito sull'importanza dell'utilizzo degli strumenti GIS nelle scienze sociali ebbe inizio nei primi anni Novanta, con un semplice scambio di idee tra scienziati interessati ai Sistemi Informativi Geografici e antropologi e sociologi. Intorno alla metà dello stesso decennio, si trasformò in un dibattito controverso. Nel 1993, alcuni dei ricercatori appartenenti ad entrambi i gruppi si incontrarono a Friday Harbour (Washington), nel tentativo di stabilire una comunicazione sistematica tra competenze diverse che consentisse il riconoscimento condiviso di problemi, in vista dell'elaborazione di soluzioni possibili.

Tra i *workshop*, va ricordato lo *Specialist Meeting*, tenutosi in Minnesota nel 1996, nel corso del quale sociologi e antropologi dichiararono che l'utilizzo canonico dei GIS, avallato da tecnici e specialisti, avrebbe finito per escluderli dal dibattito in corso e rafforzare alcuni gruppi a discapito di altri.

A fine anni Novanta, a seguito delle numerose conferenze, finanziate da organismi quali il *National Center for Geographic Information and Analysis* (NCGIA), venne prodotto il *Varenius Project*, un robusto programma di ricerca che stabilì le regole che avrebbero dovuto ispirare lo studio sull'utilizzo dei GIS e sul rapporto tra GIS e società.

Nel 2004, il report *A research Agenda for Geographic Information Science*, elaborato dai delegati dell'*University Consortium for Geographic Information Science* (UCGIS)⁴, propone una classificazione formale dei diversi gruppi di ricerca sui GIS.

Il primo potrà essere composto da quei ricercatori che sceglieranno di sposare la prospettiva della teoria sociale critica, concentrandosi su questioni quali il potere⁵, l'accesso, la sorveglianza, il modo in cui la tecnologia GIS influenzi le strutture e le pratiche sociali [McMaster e Harvey, 2010]. L'incremento di disponibilità di informazioni geo-spaziali e della conoscenza su come utilizzarle per sofisticate analisi scientifiche, lo sviluppo delle comunità scientifiche che ne fanno uso per studiare importanti problemi sociali, economici, ambientali, spaziali e relativi alla salute pubblica determina la necessità di stabilire le regole per l'utilizzo di tali dati sensibili, spaziali e non, affinché non venga lesa la *privacy*

⁴ <http://www.ucgis.org>

⁵ In origine molte geografe avevano contestato il GIS come erede dello sguardo dominante e colonialista della cartografia e come disciplina non a caso dominata in un primo momento da uomini. «In step with feminist critiques of science and geography, post structural and postcolonial authors also contend that knowledge is situated and implicated in the production of social power» [Pavlovskaya, 2007, 590]. Poi, varie applicazioni del GIS per la mappatura di ricerche sociali delle geografe femministe hanno allentato questo *divide* di genere. Ora, «geospatial technologies play an increasingly important role in the rewriting of social realities via critical epistemologies including feminism» [Pavlovskaya, 2007, 601]. Il GIS si è anche confrontato con la *queer theory* (ambito degli studi di genere che mette in discussione la rappresentazione binaria dei generi stessi) nella ricerca di due geografi che lo hanno utilizzato per una mappatura sui luoghi di ritrovo lesbo-gay a Seattle. L'idea di *queering the map* parte dunque come una provocazione, la sfida del rappresentare il non rappresentabile: «Given that the map project was guided by a queer epistemology and ontology, the project's constitutive politics necessarily reflected a process of negotiation and compromise with almost life-like forms of positivism, realism, pragmatism, and Cartesian rationality that insinuated themselves into the algorithms, hardware, and ongoing interpretation of our map production» [Brown e Knopp, 2009, 48]. I risultati positivi di esperienze di questo genere incoraggiano l'idea di insistere nella direzione di impieghi eterodossi della tecnologia.

dei soggetti coinvolti e i principi deontologici [National Research Conference, 2007].

Il secondo gruppo potrà essere composto da quei ricercatori interessati alle modalità di utilizzo dei GIS e alla gestione dei dati geografici da parte delle istituzioni. Essi, recita il report, focalizzeranno la loro attenzione sullo sviluppo di strumenti e tecniche utili a monitorare l'impatto dell'utilizzo dei GIS sui processi di *decision making* [Ibidem, 389].

Il terzo gruppo potrà essere composto dagli studiosi interessati alla prospettiva legale ed etica, alle modalità di accesso ai dati sensibili geo-spaziali, da parte dei governi, alla diffusione e vendita dei database spaziali proprietari.

Il quarto gruppo sarà composto, invece, dagli storici che si vorranno approfondire le dinamiche di evoluzione delle tecnologie informative geografiche, le dinamiche di selezione di alcuni *software* rispetto ad altri, considerando le influenze istituzionali e sociali che danno origine a tali processi selettivi.

Il quinto gruppo sarà costituito dagli studiosi interessati allo sviluppo del modello PPGIS (*Public Participation GIS*) che prevede la partecipazione pubblica nei processi di produzione dei dati georiferiti e di utilizzo dei GIS [Leitner *et al.*, 2000].

Gli incontri e i simposi citati si concludono con un auspicio: che i GIS vengano utilizzati da chi lo desidera e che non vengano controllati da gruppi ristretti e agenti di pressione, poco favorevoli alla conciliazione poiché riconoscono che nella volontà, espressa dai non esperti, di accedervi si nasconde il germe di un cambiamento e di una rivolta contro la scienza stessa, seppure lenta e mal gestita. I sociologi del territorio interessati ai GIS hanno da sempre denunciato l'abuso di autorità specialistica da parte degli esperti i quali, dal canto proprio, accentuano il conflitto perché si sentono invasi nel loro territorio. Se alcuni dei ricercatori sociali oggi affermano che i GIS sono utili alla sociologia è perché hanno avuto occasione di utilizzarli, valutarli e sperimentarne i vantaggi e di ipotizzare che la difficoltà non sta tanto nel processo di acquisizione di nuove tecniche di analisi quanto piuttosto nel dover mettere da parte capacità acquisite contribuendo, così,

a dissolvere lo *status* che li posizioni in una certa comunità piuttosto che in un'altra.

Insomma, la tecnologia GIS influisce enormemente sulla struttura della società scientifica delle scienze sociali. Essa è un derivato della scienza che ne influenza lo sviluppo⁶.

⁶ Vedi Max Weber e le relazioni tra il primo protestantesimo ascetico e il capitalismo: il primo contribuì a motivare e indirizzare le attività degli uomini verso la scienza sperimentale. La scienza, insomma, deve appoggiarsi sui valori del gruppo se intende svilupparsi. E anche un'attività così razionale come la scienza è fondata sui valori non razionali.

CAPITOLO SECONDO

GEOGRAPHIC INFORMATION SYSTEMS

I GIS aiutano il ricercatore a sintetizzare, ridurre, selezionare, discendere nei meandri della realtà sociale che può essere ricondotta alle rappresentazioni simboliche, alle linee di astrazione che caratterizzano il linguaggio della cartografia. La cartografia GIS sarà oscura, sterile, amorfa nei contenuti solo se il ricercatore sociale deciderà che il *genius loci* del vissuto debba perdersi nella forma dei grafismi. Solo se ignorerà di leggere il tempo nello spazio e di capire le direzioni localizzate dei processi sociali. È in ballo la questione della rivincita dello spazio (o *spatial turn* [Warf e Arias, 2008], che dir si voglia) che relega il dominio del tempo all'angolo delle rigidità epistemiche della diacronia e che enuncia il discorso abbozzato e provvisorio di uno spazio che diventa luogo ed esiste a partire dal momento in cui viene raccontato e rappresentato.

I GIS raccolgono e analizzano le voci e i significati, i conflitti e le contraddizioni. Non promettono la perfezione cartografica, tantomeno la bellezza di un ordine che tutto vede e fa vedere, ma permettono di sfruttare in maniera creativa l'enorme quantità di dati georiferiti a disposizione del ricercatore.

L'esigenza di rappresentare informazioni riferite al territorio e di associare ad esse attributi ad immagini è il processo fondamentale che ha dato origine all'attività di produzione delle mappe: una pratica antichissima che ha permesso all'uomo di visualizzare graficamente le informazioni spaziali e di controllare razionalmente lo spazio di suo interesse.

Le mappe sono sempre state lo strumento privilegiato di conoscenza [Zerubabel, 1992]. Esse raccontano della realtà complessa, riportandola su un piano [Monmonier, 1996]. Le mappe GIS, ancor più di quelle tradizionali, permettono all'analista di interpretare realisticamente la realtà. Tale interpretazione è facilitata dall'acquisizione di competenze [Martinotti, 2008]:

- Cartografiche⁷, le quali permettono al ricercatore di comprendere le regole utili a descrivere il territorio e interpretare i risultati dell'osservazione;
- Informatiche, le quali permettono di capire come la matrice dei dati numerici si combini con le coordinate;
- Sociologiche, le quali permettono di capire come sia stato costruito il dato e quali sono le dinamiche sociali che caratterizzano il sistema socio-spaziale rappresentato;
- Metodologiche, le quali permettono di sfruttare l'enorme flusso di informazioni geografiche per costruire modelli utili allo studio dei fenomeni sociali situati spazialmente.

2.1. Cartografia e GIS

La cartografia moderna si sviluppò tra il 1600 e il 1700. A seguito della nascita di nuove discipline quali la geodesia, la geologia e la geomorfologia, vennero elaborate le prime definizioni di ellissoide e calcolate astronomicamente le coordinate geografiche, elaborati i primi rilievi topografici e preparate le prime rappresentazioni su scale differenti [Allen, 1993].

La nascita della cartografia sociale si fa risalire, invece, al 1800. Celebri le elaborazioni cartografiche di Charles Booth [1892], il quale rappresentò la distribuzione della popolazione londinese per *status* sociale. Per la prima volta, la

⁷ La cartografia nacque nella Grecia classica, grazie al fiorire della matematica e della filosofia [Dainelli e al., 2010, 5], e si sviluppò grazie ai romani. Una delle prime rilevanti opere cartografiche è *La carta del mondo* dell'astronomo e matematico Claudio Tolomeo, realizzata nel II secolo d.C. che determina la posizione dei luoghi in base a longitudine e latitudine. Nel medioevo, la geografia tolemaica e la cartografia subirono un processo di regressione per via della dottrina cristiana che non ammetteva la sfericità terrestre. Nello stesso periodo, l'arabo Abū Abd Mhammad, conosciuto col nome di Idrisi, svolse un ruolo fondamentale nello sviluppo e diffusione della cartografia. Stabilitosi alla corte normanna di Palermo, lavorò presso Re Ruggiero II compilando informazioni geografiche ed elaborando il celebre planisfero di Idrisi, consultato dai numerosi navigatori, studiosi, astronomi, mercanti e pellegrini che si incontrarono nella città siciliana. Il Rinascimento è considerato il periodo d'oro della cartografia. Grazie all'uso di strumenti nautici, fu possibile perfezionare le proiezioni geografiche e abbandonare il sistema tolemaico. Di particolare importanza, la proiezione cilindrica del filosofo e teologo Mercatore che rappresentò i meridiani e i paralleli come delle linee perpendicolari. Inoltre, Mercatore fu in grado di risolvere il problema delle deformazioni che derivano dal processo di rappresentazione di punti di superfici sferiche su carte bidimensionali.

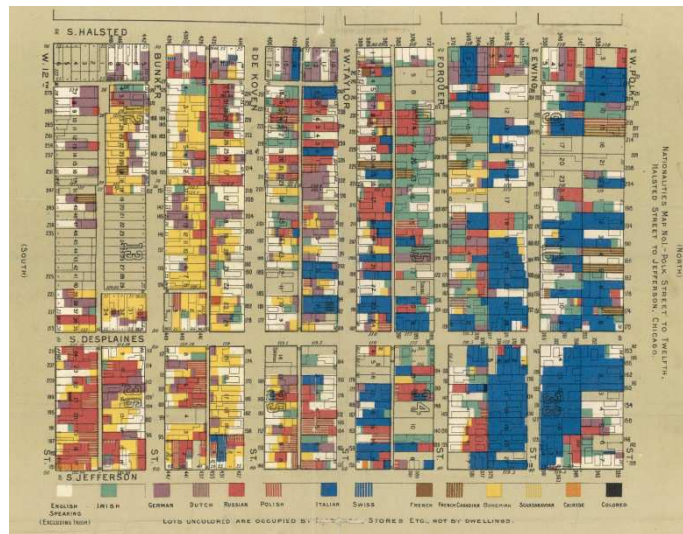
variabile socio-economica e socio-anagrafica vennero rappresentate in base alla loro distribuzione sul territorio.



(Figura 2.1. Distribuzione della popolazione londinese per status sociale. Fonte: <http://booth.lse.ac.uk>)

All’apice della Rivoluzione Industriale in Europa e Nord America, in *Hull House Maps and Papers* [Residents of Hull House, 1895], Florence Kelley raccolse una serie di mappe, prodotte seguendo l’esempio di Booth, che sono divenute un classico per l’etnografia. Per raccogliere i dati, la Kelley somministrò le interviste ai residenti della *Settlement Hull House di Chicago*⁸. Chiese loro di indicare l’area geografica entro la quale si svolgevano le loro principali attività, il numero di persone che abitavano in casa, il reddito guadagnato e il numero massimo di settimane di occupazione, nell’intero anno.

⁸ I settlement movement erano movimenti sociali riformisti che nacquero a partire dal 1880 in Inghilterra e USA. Lo scopo del movimento era quello di promuovere la convivenza pacifica tra poveri e ricchi, nello stesso quartiere. Le Settlement House nacquero nei quartieri più poveri della città. Vi lavoravano volontari appartenenti alla middle-class che contribuivano ad alleviare la povertà dei residenti del quartiere. Le case offrivano cure mediche, supporto nelle attività di scolarizzazione allo scopo di migliorare le condizioni di vita dei soggetti. La Hull House di Chicago è considerata Settlement House più importante. Fu fondata da Jane Addams e Ellen Gates Starr nel 1889.



(Figura 2.2. Florence Kelley, *Hull House Maps and Papers* [1895]. Fonte: <http://florencekelley.northwestern.edu/historical/hullhouse>)

Interessanti anche le mappe elaborate, a fine Ottocento, dalla sociologa Jane Addams che mostrano come la prostituzione, a Chicago, si concentrasse nelle zone a basso reddito [Addams, 1895]. Anticipando di quasi un secolo l'analisi multidimensionale, la Addams contribuì senz'altro allo sviluppo dell'analisi socio-spaziale.

Esclusa la parentesi temporale che va dall'inizio degli anni Cinquanta all'inizio degli anni Ottanta, in cui, con l'emergere del paradigma funzionalista in America, i sociologi del territorio cominciarono a privilegiare le *survey* e a isolare la *spatial analysis*, gli scienziati sociali hanno sempre considerato rilevanti gli sviluppi della cartografia analitica per lo studio dei fenomeni distribuiti spazialmente. La grafica spaziale veniva utilizzata, insomma, per descrivere costrutti teorici riguardanti divisioni regionali di gruppi sociali nelle città e le mappe divennero la base principale di documentazione e validazione empirica [Park *et al.*, 1925].

Gli autori dei lavori cartografici che abbiamo citato hanno sposato i principi della SISS (integrazione, modellazione esplicita spaziale, analisi *place-based*); hanno contribuito allo svecchiamento della sociologia, insistendo sulla necessità di ricorrere a strumenti computazionali complessi, anche nell'ambito della ricerca di tipo qualitativo.

Oggi, forse, la comunità dei sociologi non immagina ancora quanto gli sviluppi nell'ambito della ricerca informatica applicata alla ricerca sociale territoriale possano rivelarsi utili all'analisi. Certo è che, dall'inizio degli anni Novanta, alcuni di loro, seppur in numero esiguo, hanno cominciato ad apprezzare e utilizzare le innumerevoli applicazioni GIS per lo studio della società, delle dinamiche territoriali della città e a supporto di decisioni riguardanti il miglioramento della sua gestione.

La nostra società produce continuamente informazioni che possono essere sfruttate per la ricerca sociale. I satelliti fotografano il globo, moderne tecnologie informatiche permettono di raccogliere e gestire dati riferiti ai flussi di mobilità, alle attività e agli scambi economici. I GIS sono in grado di sfruttare tali dati, prodotti dalla società dell'informazione, analizzando la loro trasformazione nel tempo, esplorandoli alla ricerca di correlazioni ed emergenze, rappresentandoli dinamicamente, diffondendoli e condividendoli con chi desidera averne accesso, per motivi conoscitivi.

Qui, tenteremo di dare una definizione delle tecnologie che hanno aperto nuove possibilità e percorsi nel modo di pensare e studiare le complesse relazioni che si innestano nel contesto geografico.

2.2. Cosa è un GIS?

Le funzioni di un GIS possono essere spiegate a partire dall'acronimo stesso [Parker e Asencio, 2008, XIV]:

G sta per *Geographic*. I Sistemi Informativi Geografici gestiscono dati riferiti geograficamente che vengono analizzati e rappresentati attraverso le tecniche di cartografia numerica.

I sta per *Information*. Una delle potenzialità di tali sistemi è quella di immagazzinamento di una grande quantità di informazioni riferite o riferibili geograficamente, utilizzabili per scopi di analisi e ricerca diversi. Le informazioni georeferenziate legano un luogo alla presenza o all'occorrenza nello stesso di una specifica proprietà fisica (ad esempio, numero di persone, tipologia di servizi, quantità di beni, ecc.), urbanistica (ad esempio, scuole, imprese, rete viaria, ecc.);

naturalistica (ad esempio, boschi, fiumi, laghi, ecc.), critica (ad esempio, eventi naturali - quali alluvioni o terremoti - o di carattere sociale - quali crimini, manifestazioni, concerti, ecc.).

S sta per *System*. Il sistema di immagazzinamento, analisi e rappresentazione dei dati georiferiti permette di combinarli tra loro, in base ai criteri specificati dall'analista.

Un GIS, quindi, è un sistema informativo che consente di immagazzinare e gestire in modo efficiente dati geografici attraverso specifici processi di elaborazione e analisi. È un nucleo composito di *hardware*, *software*, dati, procedure e persone in grado di elaborare, analizzare e visualizzare un enorme numero di dati, nella forma di rappresentazioni cartografiche [Atzeni *et al.*, 2003].

Uno dei punti di forza di tale strumento, e ribadito dai sostenitori della SISS che se ne avvalgono, è dato dalla possibilità per l'analista di caricare nel sistema dati che tipicamente non verrebbero definiti come geografici, ma che sono portatori di informazioni georiferite, e di trasformare tali informazioni in rappresentazioni cartografiche.

2.3. Componenti di un GIS

Le componenti necessarie a condurre un'analisi GIS sono le seguenti [Ozemoy *et al.*, 1981, 92-98]:

- un pacchetto *software*, che supporti l'analista nell'implementazione delle procedure e delle funzionalità di gestione ed elaborazione dati;
- un file dati e una *basemap*, correlabili attraverso coordinate spaziali;
- un'apparecchiatura *hardware*, supporto fisico per l'archiviazione, l'elaborazione, la visualizzazione e stampa dei dati;
- un analista o, meglio, un gruppo di analisti, che posseggano competenze nel campo dell'informatica, geodetica, statistica, metodologia della ricerca sociale, sociologia, ingegneria, geografia e cartografia, in grado di utilizzare il *software* e interpretare i dati.

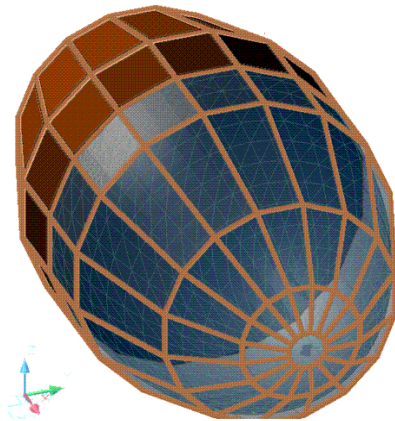
Una ricerca su Internet restituirà i *link* a decine di pacchetti *software* diversi, tra *open source* e proprietari che, in genere, comprendono una o più applicazioni utilizzabili per le diverse funzioni. Le componenti principali e necessarie del *software GIS* sono [Dainelli *et al.*, 2010, 19]:

- un database geografico che assicuri che i dati caricati e memorizzati siano organizzati e possano essere recuperati ed elaborati in qualsiasi momento;
- un'interfaccia grafica che garantisca una facile interazione analista-computer;
- una componente di inserimento e verifica dei dati, con la quale gli elementi del mondo reale possano essere archiviati all'interno del sistema;
- una componente di trasformazione e analisi dei dati che possieda tutte le funzioni e procedure destinate all'estrazione di informazione utile dai dati grezzi;
- una componente di visualizzazione e creazione di *output* che renda possibile il controllo di tutte le operazioni svolte dall'utente e la rappresentazione dei risultati di tali operazioni, a scopo di presentazione e consultazione.

2.4. Mappare la società: modelli di realtà

Per studiare e visualizzare i fenomeni situati spazialmente attraverso i GIS, occorre ricondurre lo spazio a un modello astratto composto da entità simboliche che rappresentino quelle del mondo reale. È necessario ci sia una corrispondenza biunivoca tra i punti della superficie terrestre e quelli sulla mappa, indicando la posizione dei primi rispetto ad una superficie geometrica adottata in sostituzione di quella reale [Belussi *et al.*, 2006].

Il modello matematico che rappresenta meglio la forma della Terra è quello generato dalla rotazione di un ellissoide attorno al suo asse minore, il quale riproduce lo schiacciamento dei poli e il rigonfiamento dell'equatore permettendo di calcolare con maggiore precisione angoli, distanze e coordinate [Bezoari *et al.*, 2002].

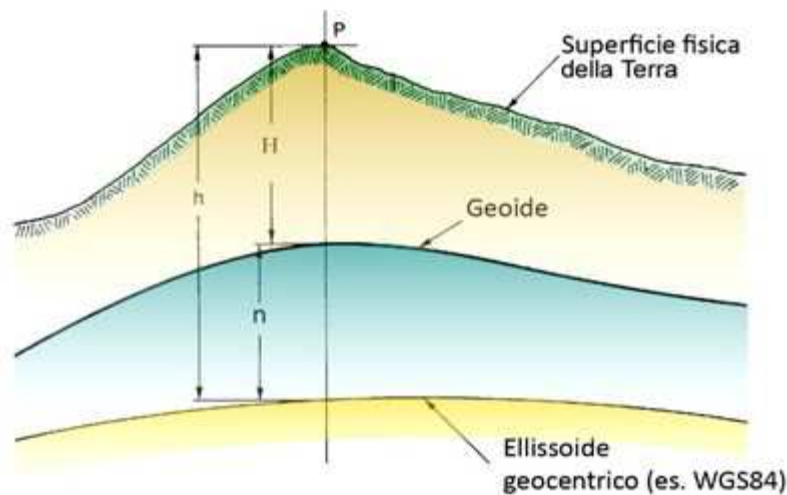


(Figura 2.3. Approssimazione poliedrica ellissoide di rotazione. Fonte: <http://assex.altervista.org>)

Il modello matematico che aderisce meglio alla realtà è il geode poich  tiene conto dell'andamento della superficie terrestre e delle sue irregolarit  [Birardi, 1988].   un modello soddisfacente a livello locale poich  consente la misurazione delle altitudini descrivendo il profilo della superficie terrestre al livello del mare. Entrambi i modelli sono indispensabili per la misurazione corretta nelle tre dimensioni [Boffi, 2004, 29]. Entrambi vengono integrati adattando le loro superfici. Le formule matematiche che permettono tale adattamento sono denominate *Datum* [Dainelli *et al.*, 2010] e ne esistono di diversi tipi.

Ad esempio, Roma40 Monte Mario   il *Datum* corrispondente al sistema geodetico nazionale italiano istituito nel 1940 e giustifica l'esigenza di approssimare meglio la superficie terrestre in un'area che comprende l'intera Italia. Oltre ai *Datum* locali, esistono numerosi *Datum* globali. Il pi  diffuso   il WGS84 (*World Geodetic System*), nato nel 1984, in seno al Dipartimento della Difesa degli Stati Uniti⁹.

⁹ In Italia, i dati geologici ufficiali sono prodotti dall'Istituto Geografico Militare.

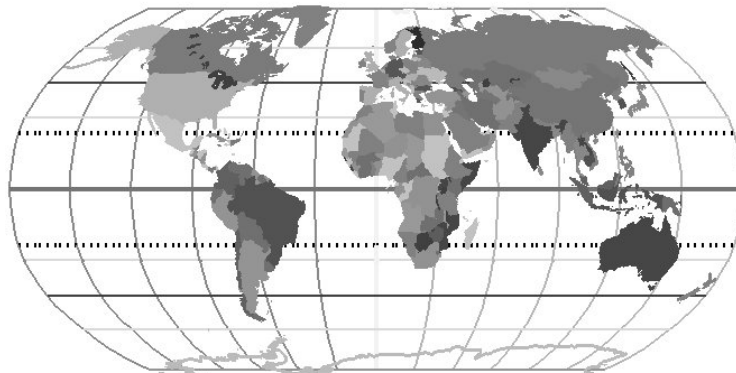


(Figura 2.4. Geoide ed ellissoide. Fonte: <http://assex.altervista.org>)

Attraverso il *Datum* geodetico, vengono calcolate le coordinate terrestri che localizzano i fenomeni, e permettono di calcolare distanza, altitudine e superficie. Le coordinate compongono la griglia che segue la superficie del modello della Terra di tipo ellissoidale [Brodeur *et al.*, 2000]. Paralleli (linee di latitudine) e meridiani (linee di longitudine) permettono di localizzare esattamente i punti sulla superficie terrestre esprimendoli in gradi, minuti e secondi di latitudine e longitudine¹⁰.

Occorre inoltre associare, al sistema di coordinate, un sistema di proiezione che permetta di riportare su un piano i punti di coordinate. Esistono proiezioni che consentono di preservare alcuni aspetti della realtà piuttosto che altri: alcune riducono le distorsioni sulle superfici, altre sugli angoli, ecc. Ad esempio, una delle proiezioni che ha lo scopo di minimizzare la distorsione dell'area e quella angolare è la Robinson (Figura 2.5.). È costruita con coordinate tabulari, invece che con formule matematiche, e rappresenta più del 75% della terra con un errore inferiore al 20% rispetto alle dimensioni in scala reale. Le direzioni sono vere lungo tutti i paralleli e lungo il meridiano centrale. La scala è vera nella fascia equatoriale [Calkins, 1996].

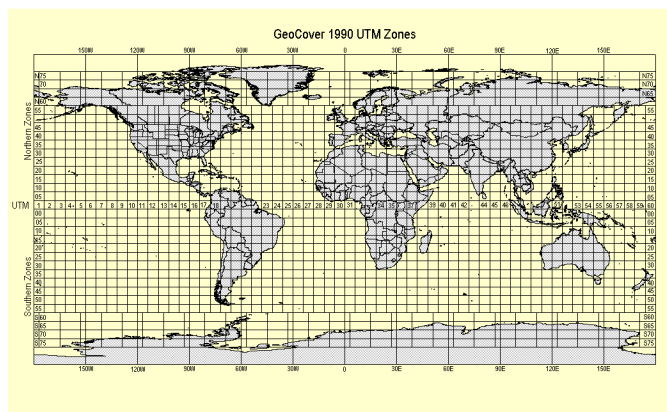
¹⁰ Il minuto è la sessantesima parte di un grado; il secondo è la sessantesima parte di un minuto.



(Figura 2.5. Proiezione Robinson. Fonte: <http://docs.bentley.com>)

Le mappe vengono costruite attraverso un processo di trasformazione, non privo di distorsioni ed errori¹¹. La scelta del più appropriato sistema di proiezione va effettuata in base agli scopi dell'analisi spaziale.

I *software* GIS permettono di scegliere tra centinaia di opzioni di sistemi di proiezione associati ai diversi *Datum*. Tra le più comuni, utilizzate per le rappresentazioni su media e grande scala (aree urbane e regionali), l'UTM (Universale Traversa di Mercatore) [Hilbert, 1960]. La griglia dei meridiani e dei paralleli è formata da linee rette che si intersecano formando angoli di 90°. L'intera superficie terrestre è suddivisa in strisce strette che si allungano a nord e a sud dell'equatore e che rappresentano deformazioni minime.

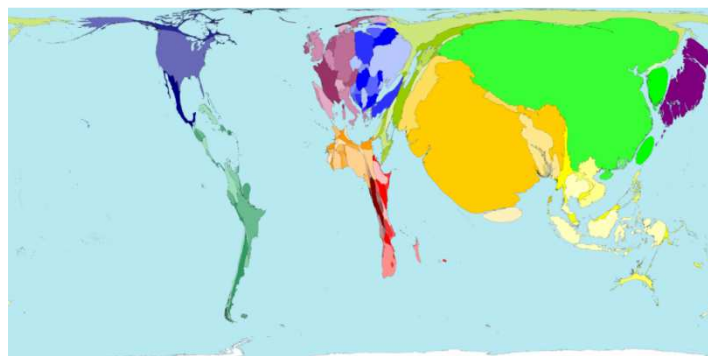


¹¹ In cartografia, la rappresentazione più comune è quella di Gauss che non presenta deformazione lungo il meridiano su cui si sceglie di centrare la rappresentazione.

(Figura 2.6. Proiezione UTM. Fonte: <http://tupian.baik.com>)

Il cartogramma umano [Szegö, 1984] è un esempio di rappresentazione non convenzionale che presenta numerose distorsioni ma aiuta a comprendere meglio le questioni oggetto di studio della sociologia, piuttosto che quelle di interesse della meteorologia o geologia, rappresentate tipicamente attraverso i cartogrammi tradizionali. La cartografia umana mostra le esperienze dei soggetti così come si esprimono nei luoghi vissuti, ridisegna i confini in base al modo in cui essi li percepiscono (*dasymeric mapping*) [Petrov, 2012].

Gastner e Newman [2004] hanno cercato di risolvere il problema della distorsione, derivante dalla dipendenza della rappresentazione dal sistema di proiezione, sviluppando un *software* che rispetta l'estensione statistica del fenomeno in esame, piuttosto che l'estensione fisica dello spazio di riferimento. I due cartografi hanno implementato e perfezionato un'applicazione derivante dalla fisica e dalla leggi di diffusione lineare dei gas. Un buon compromesso che svela distribuzioni di variabili, su scala piccola e media scala, altrimenti difficilmente intuibili. Considerando, ad esempio, la densità della popolazione mondiale di suicidi nel 2002, questa viene idealmente fatta fluire verso le aree a maggiore densità, espandendo e contraendo le superfici¹².



(Figura 2.8. Cartogramma *Worldmapper* costruito attraverso il metodo Gastner e Newman. Il Nord Africa è la regione caratterizzata dal più basso tasso di suicidi in rapporto alla popolazione nordafricana. La superficie si espande verso il Giappone in cui si registrano 245 morti su un milione di persone. Fonte: <http://worldmapper.org>)

¹² Il plugin ArcGIS che permette di utilizzare l'algoritmo Gastner e Newman e di costruire cartogrammi seguendo il metodo della diffusione è scaricabile dalla pagina dell'Imperial College di Londra 'Complexity and Networks' <http://www2.imperial.ac.uk/~mgastner/>

Numerose cartografie, realizzate seguendo il metodo della diffusione di Gastner e Newman, possono essere consultate e scaricate dal sito *Worldmapper - the world as you've never seen before*¹³, un progetto che nasce dalla collaborazione tra i ricercatori, esperti in Sociologia, Geografia e Medicina Sociale, del *Social and Spatial Inequalities Research Group* dell'Università di Sheffield (UK) e del *Center for the Study of Complex Systems* dell'Università del Michigan (USA).

2.5. Questione di scala

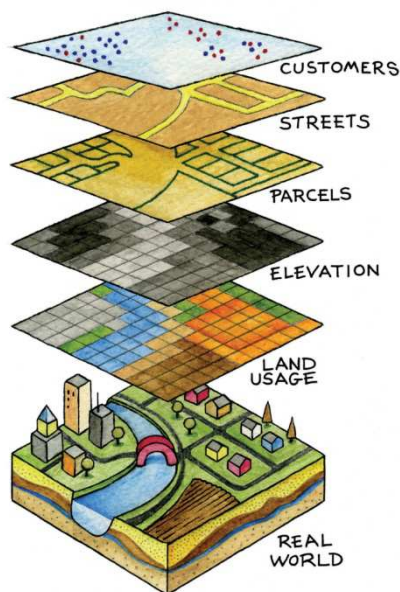
È il ricercatore a decidere quanto la distorsione di cui è portatrice una proiezione sia accettabile. È lui a decidere quale scala utilizzare. Mappare la distribuzione dei residenti di una città densamente popolata usando una mappa a piccola scala significa rendere virtualmente invisibile la distribuzione stessa [McCoy, 2004]. Di contro, scegliere di mappare su grande scala una vasta zona rurale scarsamente popolata produce l'effetto inverso. È attraverso i sistemi di proiezione che è possibile costruire mappe che rappresentino il mondo reale in versione ridotta.

La scala, espressa sottoforma di frazione numerica, in cui al numeratore si trova la dimensione sulla mappa e, al denominatore, la corrispondente dimensione del mondo reale, riduce le aree in base alle esigenze di analisi calcolando il rapporto tra la dimensione di un oggetto e la sua dimensione reale [Monti e Pinto, 2002].

¹³ <http://worldmapper.org>;

2.6. Modelli cartografici

Per catturare la complessità sociale, l'analista GIS utilizza i *layer*¹⁴, strati informativi che possono essere sovrapposti l'uno sull'altro rappresentando ciascuno un'unica classe di oggetti.



(Figura 2.11. Layer. Fonte: <http://catalog.flatworldknowledge.com>)

I *layer* permettono all'analista di capitalizzare il lavoro di ricerca: possono essere riutilizzati nel corso di studi futuri, condivisi con altri ricercatori che ne arricchiranno, se vorranno, il contenuto [Schiavi, 2002]. L'operazione di sovrapposizione di più *layer* permette di concettualizzare il mondo sociale, di integrare le informazioni spaziali e le rappresentazioni delle proprietà multiple dei luoghi. Ogni strato geografico è rappresentativo di informazioni diverse, accomunate dallo stesso sistema di riferimento. La sovrapposizione dei *layer* è, dunque, possibile perché essi sono riferiti a precise porzioni della sfera terrestre e misurati su noti sistemi di riferimento locali o globali.

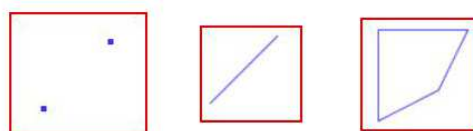
¹⁴ Prima della diffusione dei moderni software GIS, il tradizionale procedimento fisico di composizione delle mappe avveniva usando fogli trasparenti che venivano sovrapposti l'uno sull'altro.

Trasformando i fenomeni del mondo reale in oggetti riconducibili a forme geometriche localizzabili e dotate di attributi, sarà possibile specificare quelle proprietà qualitative, quantitative e spaziali che facilitano l'interpretazione dei processi sociali [Travaglini, 2004].

Esistono due modelli tra loro complementari, utili a rappresentare i dati nello spazio: vettoriale e *raster*. Il primo è di tipo discreto, il secondo, continuo. Entrambi, per la loro diversa natura, concettualizzano diversamente i fenomeni. A ciascuno dei due modelli, si riconducono tecniche diverse di gestione dei dati.

2.7. People manipulate object but live in fields. Modello vettoriale e raster

In base al modello vettoriale, ogni dato può essere rappresentato sotto forma di primitiva geografica (*punto*, *linea* e *poligono*) [Schiavi, 2002].



(Figura 2.12. Oggetti base utilizzati nei GIS: punto, linea, poligono)

Il *punto* localizza l'oggetto nello spazio. Si definisce *nodo* nel momento in cui è collegato ad un altro punto. La *linea* collega coppie di punti e rappresenta posizione, direzione e lunghezza delle entità del mondo reale. Il *poligono*, o area, definisce, oltre alla posizione, anche la superficie discreta di un'area. Ad ogni primitiva, può essere sempre associata una tabella alfanumerica in cui vengono conservate le informazioni riguardanti l'entità o oggetto [Travaglini, 2004].

Ad esempio, in una rappresentazione puntuale delle aziende confiscate alla mafia in Sicilia, ad ogni punto potrebbe essere associato un valore relativo alla data di confisca, al numero di dipendenti, al fatturato, alla ragione sociale, ecc. Dal punto di vista della semplice rappresentazione, la possibilità di sfruttare una tabella di attributi da associare agli elementi geografici permette di realizzare più cartografie tematiche, evidenziando, di volta in volta, i dati che ci interessano e che sono archiviati in tabella.

Si potrebbe, ad esempio, scegliere di rappresentare le aziende confiscate in base al tipo di attività. Ma, al di là della semplice rappresentazione, la tabella associata si rivela utile per effettuare interrogazioni o *query* al database. Si potrà rispondere a domande quali: «Dove si trovano e quali sono le aziende confiscate appartenenti al settore delle costruzioni?»; «Dove si trovano e quali sono le aziende o i beni confiscati alla stessa famiglia?».

Entità fisiche, amministrative, produttive o eventi, che sono circoscrivibili in uno spazio definito, saranno scomponibili concettualmente e spazialmente, rappresentabili in oggetti geometrici [Boffi, 2004, 47]. E anche i fenomeni caratterizzati da una natura fisica continua potranno essere rappresentati come oggetti spaziali, frazionandoli e delimitando le aree omogenee al fine di costituire un unico oggetto.

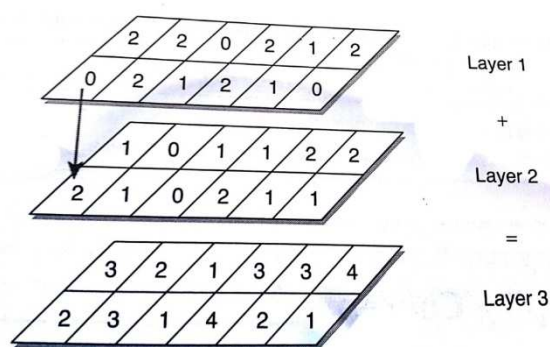
Queste sono le principali ragioni per cui gli analisti GIS definiscono l'intero processo di analisi e gestione dei dati georiferiti non come semplici rappresentazioni geografiche bensì come elaborazioni di veri e propri geodatabase informatizzati.

Proprio in virtù della possibilità di georiferire gli oggetti attraverso il modello vettoriale, è possibile verificare quale sia il rapporto geometrico-spaziale che lega le entità e le forme. Alla domanda: «Quali beni e aziende confiscati ai fratelli Graviano si trovano nel comune di Palermo?», il GIS risponde effettuando una ricerca del dato puntuale delle aziende confiscate in quello poligonale del comune di Palermo.

I dati in formato *raster* vengono, invece, memorizzati attraverso una griglia regolare composta da celle alle quali viene assegnato un valore alfanumerico che rappresenta una proprietà dello spazio [Monti e Pinto, 2002]. Ogni cella corrisponde, nella realtà, a una estensione di territorio. Ad esempio, un parco può essere rappresentato da insiemi di celle adiacenti alle quali viene assegnato uno stesso valore.

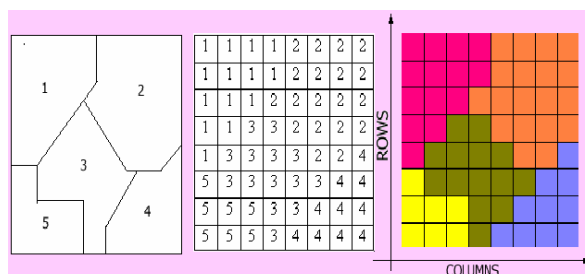
Le celle dei modelli *raster* sono spesso rappresentate come quadrati di dimensione fissa nei quali vengono codificati i dati: un attributo per ogni cella [ESRI, 2009]. La Figura 2.13 mostra come il modello *raster* possa essere utile per conteggiare il

numero di residenti in un certo quartiere sovrapponendo *layer* portatori ognuno di un attributo diverso. Il *layer 1* potrebbe rappresentare, ad esempio, il numero di adulti a partire dai 18 anni, il *layer 2*, il numero di bambini. Oltre alla somma, il modello *raster* permette di compiere diverse operazioni matematiche, a cominciare dalle aritmetiche [Steinberg e Steinberg, 2006]. In questo modo, è possibile implementare tecniche statistiche avanzate per l'analisi dei dati.



(Figura 2.13. Codifica layer raster. Fonte: Steinberg e Steinberg, 2006, 25)

Sebbene i due modelli siano da considerarsi complementari, quello vettoriale è preferibile nel caso in cui si abbia necessità di analizzare e rappresentare dati topologici riferiti dettagliati. Dall'altro lato, il vantaggio del modello *raster* è dato dal fatto di poter manipolare matematicamente differenti *layer* assegnando ad ogni cella un codice ed elaborando un sistema di codifica in base alle esigenze di ricerca [Bezoari *et al.*, 2002].



(Figura 2.14. Modelli *raster* e vettoriale)

Il modello dati vettoriale è più flessibile poiché si aggancia anche a informazioni alfanumeriche descrittive [Perencsik *et al.*, 2004]. Viene, infatti, utilizzato nell'ambito della geografia amministrativa, della suddivisione catastale, della viabilità. Il modello *raster*, invece, è più approssimativo nel disegno delle geometrie, poco flessibile nel collegamento a dati descrittivi ma estremamente adatto a compiere operazioni computazionali. Facendo uso dell'algebra basata sulla combinazione di operatori matematici, logici e aritmetici consente di creare modelli della realtà in modo rapido ed efficiente. In generale, è preferibile usare il modello vettoriale quando è necessaria la precisione delle forme geometriche e delle relazioni spaziali. Si preferisce il modello dei dati *raster* quando si è interessati all'analisi rapida dei fenomeni e alla creazione di modelli [Ticca, 2009, 179]. Spesso, come vedremo successivamente, è indispensabile analizzare i dati con entrambi i tipi di modelli.

2.7.1. Formati digitali di codifica dell'informazione geografica

L'informazione geografica digitale può essere archiviata secondo diversi formati di codifica. La scelta del formato dipende soprattutto dal tipo di modello spaziale attraverso il quale l'informazione è strutturata (*raster* o vettore). ESRI, la casa produttrice del software GIS più diffuso (ArcGIS) ha prodotto degli standard *de facto* di formati di memorizzazione dei dati geografici. Sono formati flessibili, resi pubblici e utilizzabili attraverso tutti i tipi di software GIS [ESRI, 2009].

Il formato di archiviazione dei dati vettoriali è lo *shapefile*. Permette di registrare localizzazione, forma e attributi delle entità spaziali attraverso le tre principali classi di oggetti: punti, linee, poligoni. Uno *shapefile* è portatore sia di informazione geometrica che tabellare. È composto da un set di file che comprende lo *shape.shp*, la *tabella.shx* (che permette di ricostruire il raccordo tra geometria e informazione tabellare) e ulteriori datafile accessori come gli indici di attributo (*file.ain*, *file.aih*), i sistemi di riferimento (*file.prj*) e i metadata (*file.xml*) [ESRI Italia, 2013; Kennedy e Meyers, 1977].

I formati delle immagini, invece, sono archiviabili in un GIS attraverso una griglia *raster* georeferenziata; una griglia, cioè, di cui è registrata la posizione di ogni *pixel* rispetto a un sistema di riferimento [Steinberg e Steinberg, 2006]. Esempi di immagini archiviabili in un GIS sono le rilevazioni satellitari, fotografie, scansioni di fotografie. Il dataset di un'immagine georeferenziata contiene sempre il file immagine registrato in diversi formati (come vedremo nel corso del paragrafo successivo, il più comune è il *.tiff*) [McCoy, 2004]. Oltre al file immagine, è possibile ritrovare nel *dataset* ulteriori file accessori che registrano informazioni per velocizzare la visualizzazione (*file.aux*, *file.rrd*) e, anche in questo caso, metadata (*file.xml*)¹⁵.

2.8. Georeferenziazione con Quantum GIS

Un GIS gode della capacità di georeferenziare, di attribuire cioè ad ogni entità rappresentata le sue coordinate spaziali reali, in base al sistema di riferimento scelto [Date, 2003].

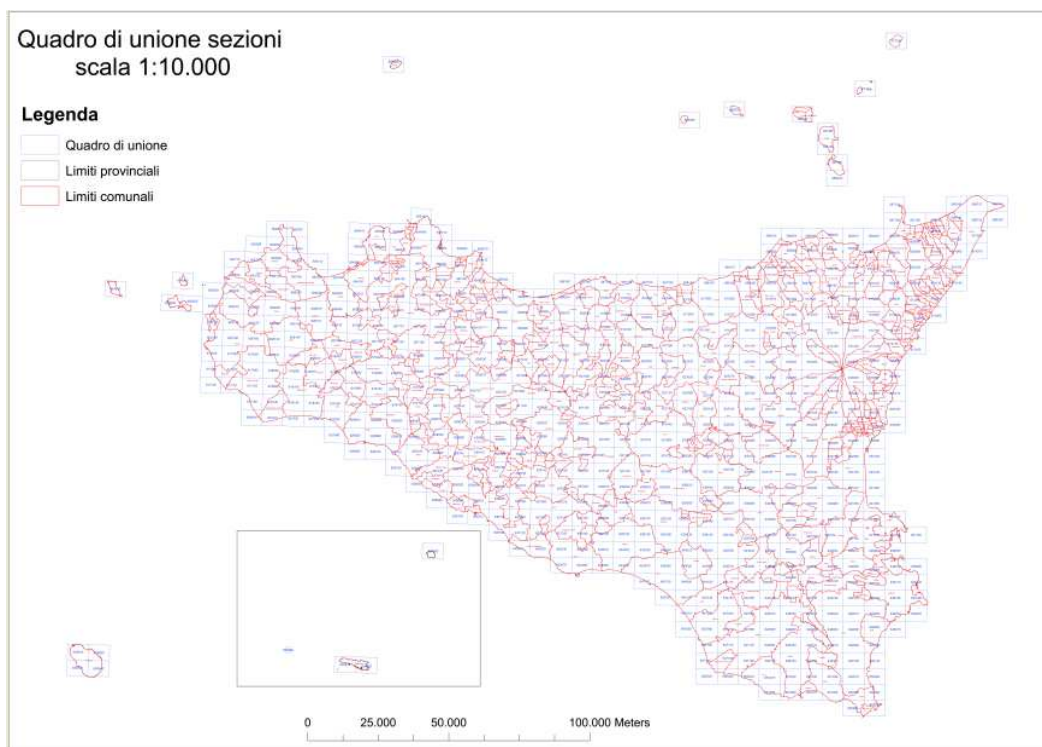
Per capire come utilizzare i dati *raster* e vettoriali, riportiamo un esempio di georeferenziazione. Per la sua esecuzione, si è scelto di utilizzare il software *open source* QGIS [Steinberg e Steinber, 2006].

È possibile accedere ai dati cartografici utili all'esercitazione, dal sito del Dipartimento Urbanistica della Regione Sicilia (www.sitr.regione.sicilia.it). Il file in formato *raster* che verrà utilizzato nel corso della georeferenziazione è il CTR denominato *f594040*¹⁶ ed è in formato *.tiff* o *.tif* (*Tagged Image File Format*)¹⁷.

¹⁵ Si discuterà dei Metadata nel corso del Capitolo Quarto.

¹⁶ Per la conoscenza e la rappresentazione del proprio territorio, la Regione Siciliana ha scelto di dotarsi della Carta Tecnica Regionale alla scala 1:10.000 come cartografia di base a copertura dell'intero territorio regionale. La Carta Tecnica Regionale in scala 1:10000 è rappresentata nella proiezione di Gauss. Essendo la Sicilia compresa tra 36° latitudine Sud e 38° latitudine Nord, ogni sezione è contraddistinta da un nome di località e da un numero di sei cifre di cui le prime tre indicano il numero del foglio in scala 1:50.000 in cui la sezione ricade, le successive due cifre indicano in quale posizione si trova all'interno del foglio (da 01 a 16) e l'ultima cifra è zero. Il quadro d'unione è l'insieme dei fogli catastali di un intero comune, provincia o regione.

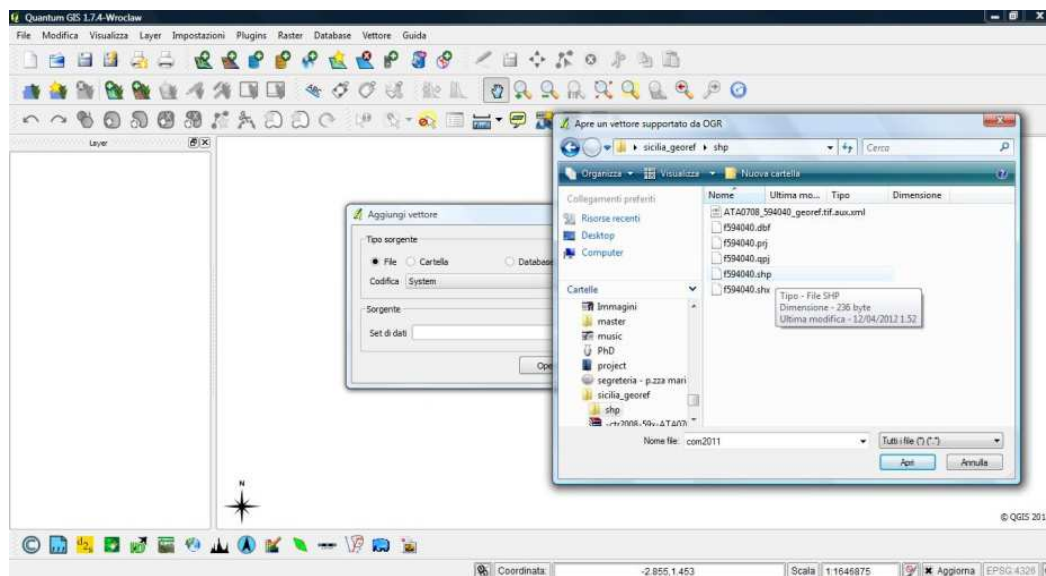
¹⁷ Il *.tiff* o *.tif* è uno dei formati raster più diffusi. La conversione di una qualsiasi immagine o di un file *.pdf* in tale formato è operabile attraverso un software di elaborazione di immagini digitali, come Adobe Photoshop o GIMP.



(Figura 2.15. Snapshot QGIS. CTR in formato .tiff)

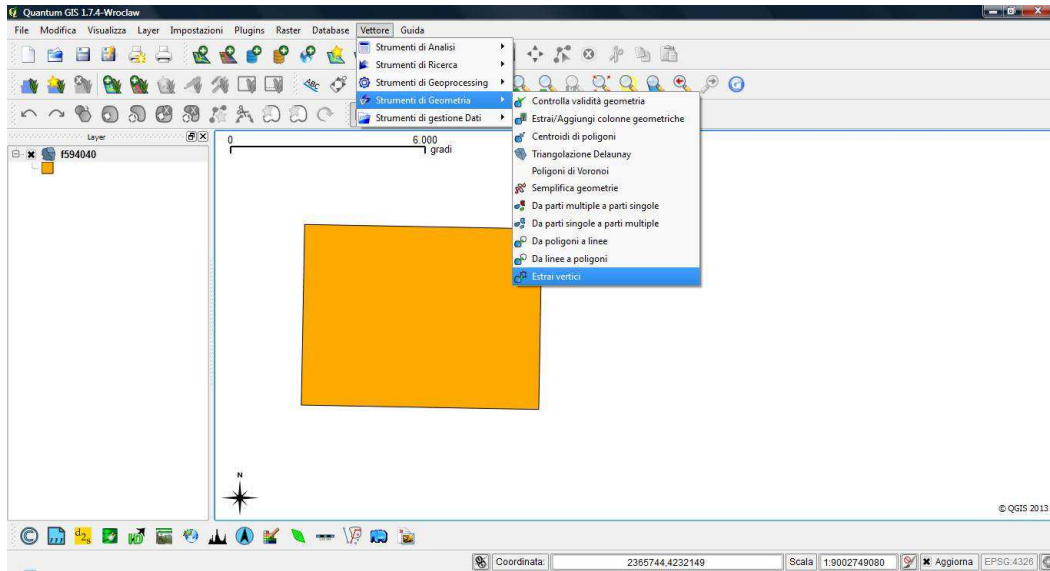
La georeferenziazione avviene individuando le coordinate dei punti che coincidono coi quattro angoli della mappa .tiff.

Si apra il software QGIS e si aggiunga il foglio vettore *f594040.shp*.



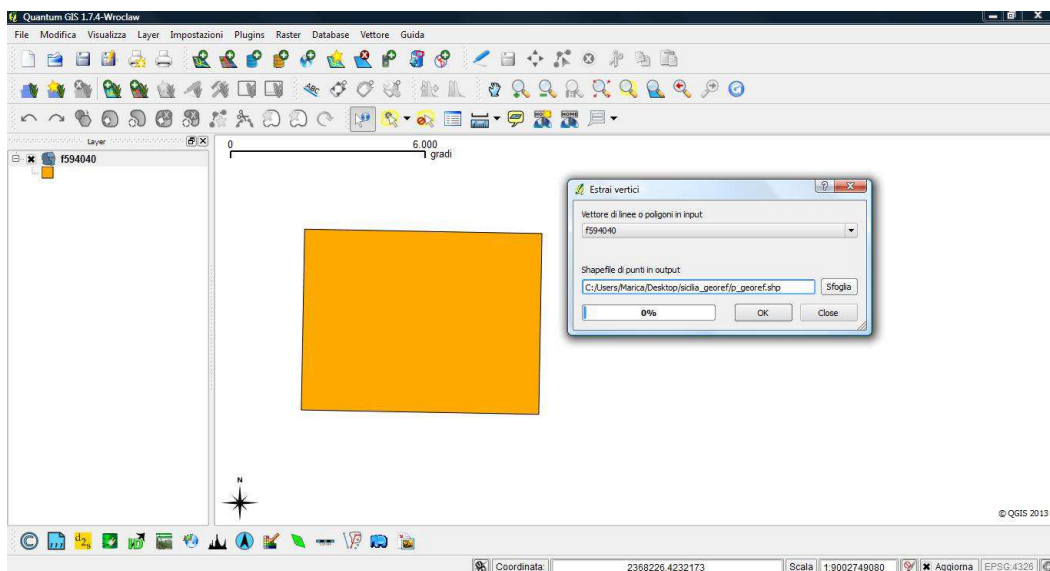
(Figura 2.16. Snapshot QGIS. Aggiunta vettore *f594040.shp*)

Dallo strumento *Vettore*, si sceglie l'opzione *Strumenti di Geometria / Estrai Vertici*.



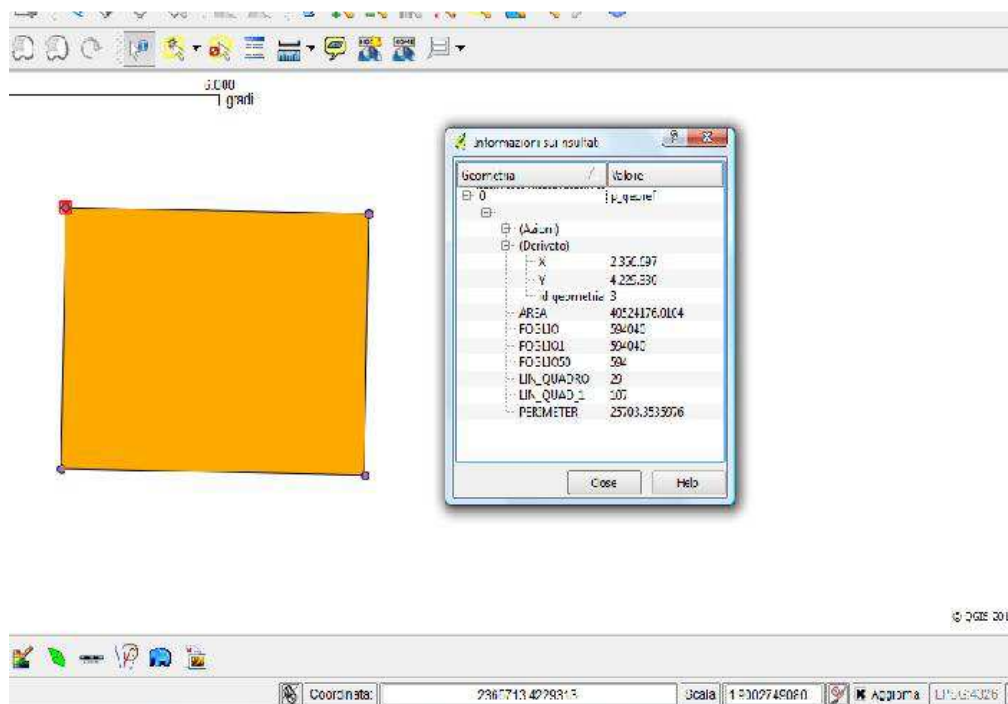
(Figura 2.17. Snapshot QGIS. Estrazione vertici)

Si denomini il file di punti in output *p_georef.shp*.



(Figura 2.18. Snapshot QGIS. Denominazione file di punti .shp in uscita)

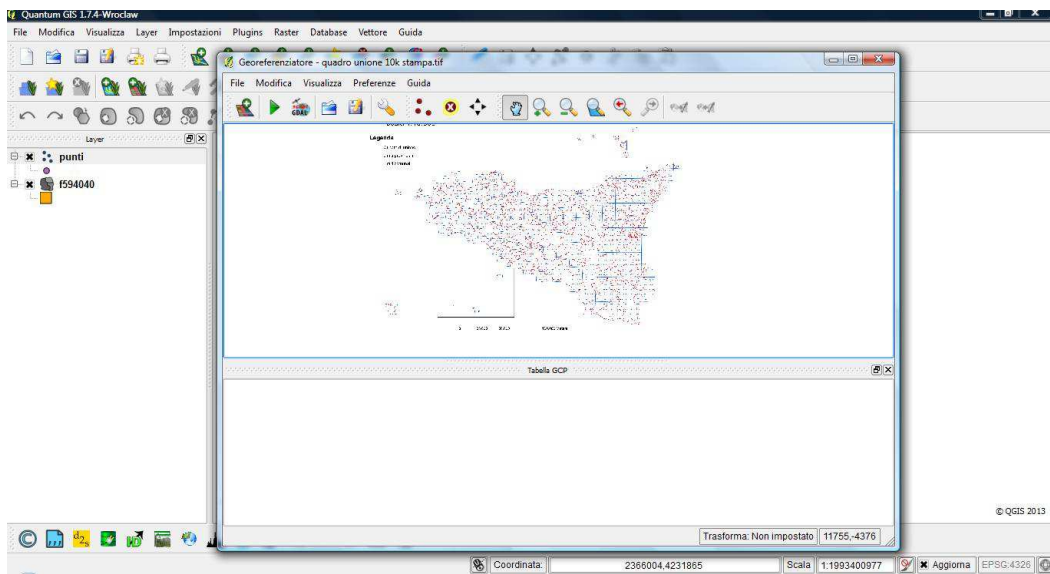
Dopo aver cliccato sul comando *Informazioni elementi*, sarà possibile ottenere le informazioni relative ad ognuno dei quattro vertici del foglio, selezionandoli.



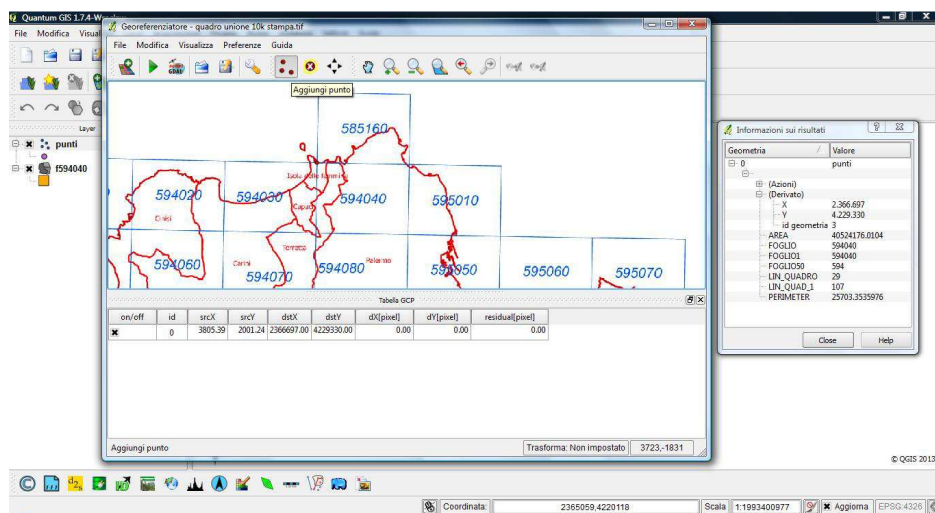
(Figura 2.19. Snapshot QGIS. Informazioni elementi)

Da *Geometria Derivato*, sarà possibile leggere il valore delle coordinate.

A questo punto, occorrerà scegliere dal *Menu Plugins*, il *Georeferenziatore*. Sarà possibile aggiungere il file *raster* in formato *.tiff* e iniziare la georeferenziazione ingrandendo il file *raster* in modo da poter visualizzare chiaramente i vertici del foglio *f594040*, selezionandoli uno per volta e caricando, nella finestra di inserimento delle coordinate, quelle ottenute da *Informazioni Elementi*.

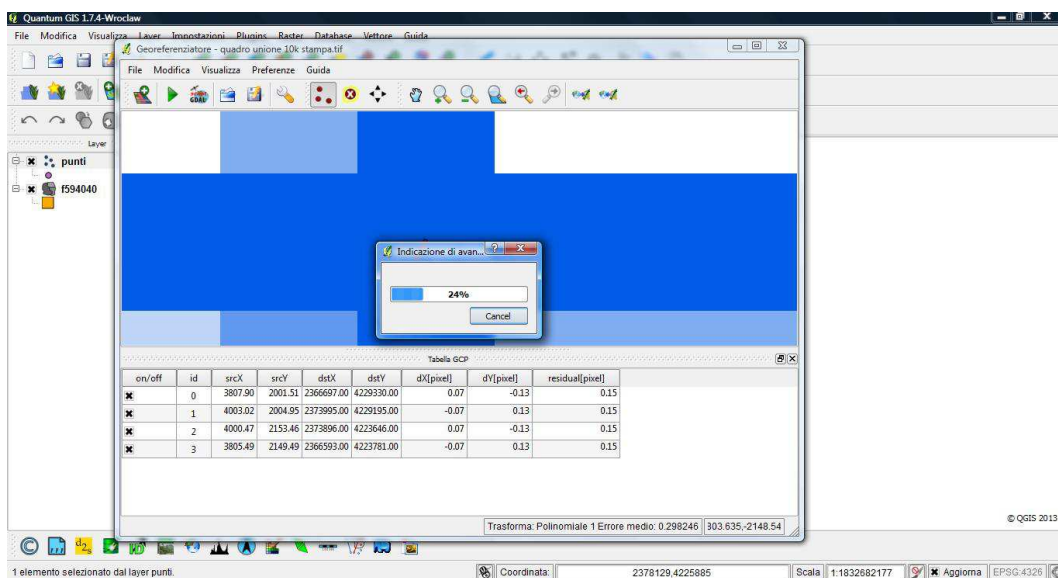


(Figura 2.20. Snapshot QGIS. Georeferenziatore)



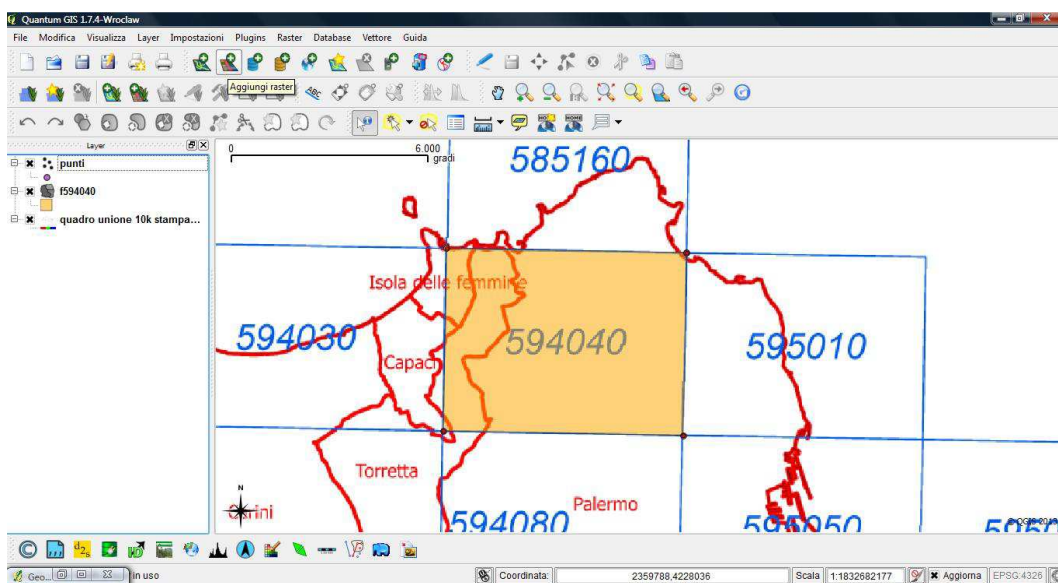
(Figura 2.21. Snapshot QGIS. Georeferenziatore)

Cliccando sul tasto *Play*, si dia il via al processo di georeferenziazione alla fine del quale sarà possibile salvare il file *raster* georeferenziato da aggiungere tra i *layer* di QGIS.

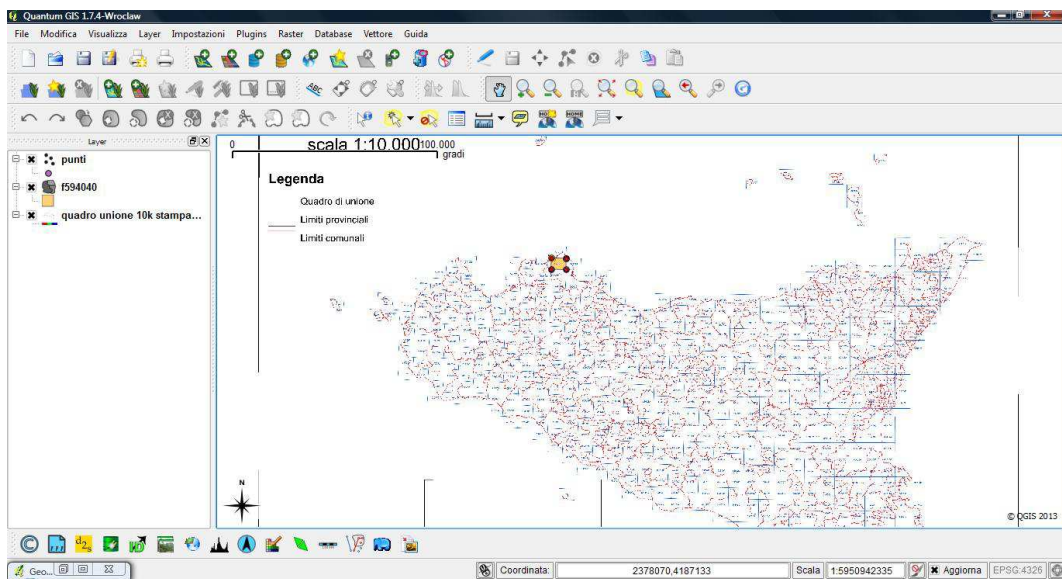


(Figura 2.22. Snapshot QGIS. Georeferenziatore)

Si noterà come l'immagine *raster* georeferenziata aderirà perfettamente al foglio vettore posizionato in base al sistema di riferimento scelto.



(Figura 2.23. Snapshot QGIS. Aggiunta del raster georeferenziato)



(Figura 2.24. Snapshot QGIS. Raster georeferenziato)

Come si vedrà nel corso del Capitolo Quinto dedicato al *Qualitative GIS Approach*, per tutti quei ricercatori sociali che hanno previsto di trattare dati qualitativi attraverso gli strumenti GIS, sarà indispensabile conoscere le procedure di georeferenziazione appena descritte. La georeferenziazione permette, infatti, di archiviare, organizzare, analizzare e rappresentare dati qualitativi e quantitativi georiferibili e di intraprendere un percorso di ricerca iterativo tra teoria, dati e interpretazione.

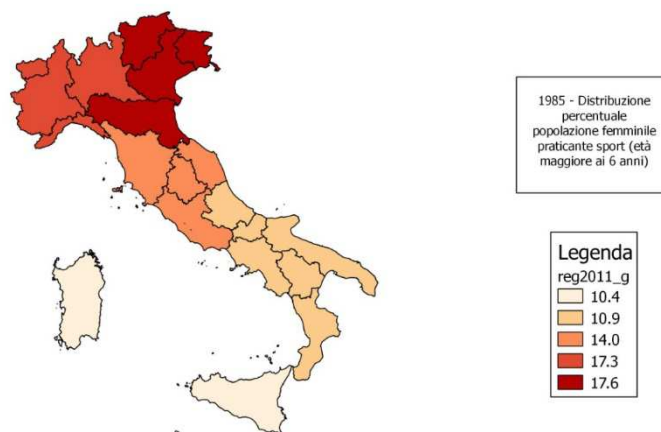
2.9. Operativizzazione dei concetti nella GIS analysis: entità e attributi

L'*attributo* è la caratteristica di un determinato oggetto o *entità* localizzabile [Fowlerm, 2004]. Attraverso il disegno della ricerca (ossia il piano di lavoro che stabilisce quali, quanti oggetti verranno osservati e dove, con quale criterio saranno scelti, con quali strumenti le loro caratteristiche verranno registrate), l'analista traduce in termini empirici gli attributi in variabili [Boffi, 2004]. Occorre, prima di tutto, che scelga il tipo di oggetto, ossia l'unità di analisi spaziale. È possibile che si tratti di aggregati territoriali (ad esempio, quartieri, comuni, province, regioni), o di manifestazioni culturali, eventi (ad esempio, conflitti armati, consultazioni elettorali) [Intesa GIS, 2004]. Non esiste limite a ciò

che può essere definito unità di ricerca empirica spaziale, a patto che sia riferibile geograficamente. La scelta dell'unità, entità o oggetto varia in base agli interessi del ricercatore e alle questioni cognitive che intende porsi.

Oltre a scegliere l'unità, questi dovrà individuare l'ambito spazio-temporale entro il quale intende osservarla. Potrà, così, determinare la popolazione di riferimento e, quindi, l'insieme dei casi della ricerca.

Ad esempio, se l'unità è il praticante sport di genere femminile e di età maggiore ai 6 anni e l'ambito spazio-temporale è l'Italia nel 1985, la popolazione di riferimento sarà costituita dall'insieme delle donne residenti in Italia che nel 1985 praticavano almeno uno sport.



(Figura 2.25. Distribuzione percentuale popolazione femminile praticante sport – età maggiore ai 6 anni. Nostra elaborazione con QGIS. Indagine su Sport e tempo libero)

Gli attributi assumono, sulle unità o entità spaziali alle quali si riferiscono, stati diversi che variano da entità a entità. Per tradurre un attributo in termini empirici, occorre darne una definizione operativa e stabilire, quindi, le procedure per rilevare gli stati della proprietà [Corbetta *et al.*, 2001, 21]. L'applicazione della definizione operativa consente di trasformare una proprietà in variabile e i suoi stati in modalità della variabile stessa. Per facilitare la registrazione di tali modalità e la conseguente gestione delle stesse nel GIS, viene assegnata ad ognuna un diverso valore simbolico che viene registrato nel database GIS [Pfaff, 2004]. La decisione su come operativizzare un attributo o proprietà deriva dalla

sensibilità del ricercatore che dovrà opportunamente esplicitare le sue scelte [Marradi, 1987, 25]. Attraverso la GIS analysis, la presunta soggettività di tale scelta si riduce poiché la comunità scientifica di riferimento è spinta a produrre definizioni operative convenzionali di carattere intersoggettivo che possano essere sperimentate dai diversi studiosi proprio attraverso i sistemi informativi geografici.

Le variabili nascono dal processo di operativizzazione degli attributi spaziali. Esse si classificano in base al tipo di operazioni logico-matematiche alle quali possono essere sottoposte. Esistono diverse tipologie attraverso le quali classificarle. La differenza tra una tipologia e un'altra è data dall'operazione intellettuale che anima il processo di operativizzazione, il quale dipende, a sua volta, dal tipo di attributo che si intende operativizzare [Corbetta, 2001, 28-29]:

- Variabili categoriali. Nascono dall'operativizzazione di attributi che assumono stati discreti non ordinabili. Questi ultimi sono descritti in modo nominalistico (con il nome di un comune o il suo codice identificativo). Inoltre, possono indicare le classi a cui appartengono gli oggetti (edifici ad uso abitativo, ad uso industriale; individui di genere maschile o femminile). Esistono, insomma, confini molto netti tra gli stati: non è possibile immaginare stati intermedi. La natura non ordinabile degli stessi non facilita il rinvenimento di una relazione d'ordine. Ciò significa che non è possibile ordinarli in base a un qualche criterio ma stabilire solamente delle relazioni logiche: un caso può essere uguale o diverso da un altro [Chrisman, 1998]. Il GIS potrà solo classificare e individuare, cioè, un insieme di classi che corrispondono agli stati che un attributo può assumere. Per semplicità, nel processo di codifica e di inserimento dati nel database, possiamo decidere di assegnare un codice numerico ai dati nominali. Ad esempio, si può decidere di considerare le localizzazioni urbana, suburbana o rurale quali attributi della variabile localizzazione. I numeri assegnati (1 = città; 2 = suburbana; 3 = rurale) saranno delle semplici etichette nominali.

- Variabili ordinali. Misurano delle grandezze ma indicano solo la posizione relativa delle stesse, su una scala. Ad esempio, i centri urbani sono spesso classificati nelle mappe facendo riferimento alla popolazione residente e sono raggruppati in classi ordinali (grandi, medi, piccoli centri urbani). L'esistenza di un ordine sotteso agli stati permette di rilevare non solo relazioni di eguaglianza e diseguaglianza ma anche di ordine, cioè «maggiore di» e «minore di». La definizione operativa si basa sull'ordinamento; pertanto, sebbene non sia nota la distanza che intercorre tra i diversi stati, l'attribuzione dei loro valori non potrà essere arbitraria, ma dovrà rispettare l'ordine stesso. Gli attributi non possono essere sommati, sottratti, divisi (ad esempio, la somma di due entità spaziali piccole non dà come risultato un oggetto più grande). In una GIS *analysis*, è opportuno codificare i dati ordinali attraverso codici numerici piuttosto che attraverso stringhe nominali, di modo che possano essere inseriti in modelli matematici che calcolino le tre relazioni d'ordine.
- Variabili cardinali, a intervalli e a rapporti. Misurano, con valori continui, l'intensità di un attributo con un certo grado di precisione (ad esempio, età di una persona, altezza di un edificio, reddito di una famiglia). I valori assegnati agli stati hanno un pieno significato numerico. I numeri non possiedono solo le caratteristiche ordinali ma anche quelle cardinali. Pertanto, non solo il GIS calcolerà relazioni di eguaglianza e differenza o relazioni d'ordine ma potrà effettuare anche tutte le operazioni aritmetiche sui valori. Calcolerà le distanze fra gli stati in virtù dell'esistenza di un'unità di riferimento e sarà in grado di sottoporre le variabili a differenti elaborazioni statistiche. Gli attributi misurati a intervalli non hanno una base assoluta o naturale di riferimento ma solo una base convenzionale (ad esempio, la scala delle temperature comunemente in uso). In un'analisi GIS, gli attributi misurati ad intervallo offrono all'analista la possibilità di condurre un'analisi quantitativa dettagliata poiché le operazioni statisticomatematiche calcoleranno le distanze tra intervalli di valori, anche geografici. Le misurazioni a rapporti fanno, invece, riferimento a una base

assoluta (ad esempio, l'anno di nascita è la base assoluta per misurare l'età).

Le definizioni operative che generano le variabili cardinali sono la misurazione e il conteggio [*Ibidem*]. Nel primo caso, l'attributo da misurare è continuo¹⁸ e la comunità scientifica ha elaborato un'unità di misura condivisa di riferimento con la quale confrontare la grandezza da misurare. Nel secondo caso, l'attributo da registrare è discreto¹⁹ e l'operativizzazione consiste nel contare gli elementi posseduti dall'entità spaziale con i quali questa è in relazione (ad esempio, il numero di addetti di un'impresa confiscata alla mafia). Gli attributi misurati a rapporti hanno un'unità di conto naturale: essa non è stata inventata. Perciò, intorno al suo uso, non è necessario che la comunità scientifica stabilisca un accordo.

La struttura del *geodatabase*, di cui si discuterà nel corso del Capitolo Terzo, permette di organizzare i dati collegandoli alla loro collocazione spaziale e di registrare gli attributi di una variabile, sia in forma numerica che non numerica [Date, 2003; Fowlerm, 2004; Goramasca, 2004; Dainelli, 2010].

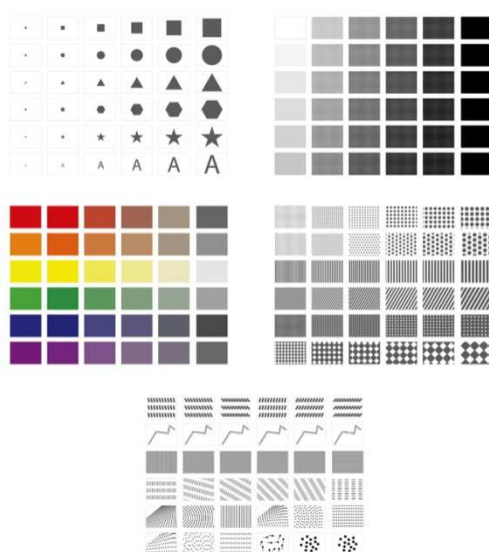
Ad esempio, se siamo interessati a definire il concetto di benessere in una comunità rurale, potremmo considerare come indicatori di quella variabile la disponibilità di servizi medico-sanitari, chiederci quali tipi di servizi siano disponibili, se esistano cliniche o ospedali, in quali e quante ore tali servizi siano disponibili, se siano localizzati nella zona rurale di nostro interesse o nell'area urbana, quanti posti letto siano disponibili, ecc. Gli attributi della variabile "servizi medico-sanitari" potrebbero essere registrati nel *geodatabase* con le etichette *tipo, ore, localizzazione, letti*, ecc. e utilizzando stringhe o numeri in base alla natura degli attributi delle variabili in questione, tra nominale, ordinale e cardinale [Steinberg e Steinberg, 2006].

Come scritto, i GIS riescono a tradurre i dati georiferiti in simboli. Bertin [1967; 1977] distingue due famiglie di variabili visive: sei variabili retiniche

¹⁸ Un attributo si dice continuo se assume infiniti stati intermedi in un dato intervallo fra due stati qualsiasi [Corbetta, 2001, 29].

¹⁹ Un attributo si dice discreto se assume stati finiti, non frazionabili [Corbetta, 2001, 30].

(forma, densità cromatica, orientamento, retinatura, colore, dimensione) e due di posizione, in base agli assi cartesiani (nel caso dei grafici) e alle coordinate (nel caso delle mappe). Le variabili di posizione fissano un simbolo sul piano; quelle retiniche lo rappresentano graficamente descrivendo relazioni di somiglianza, ordine, proporzionalità.



(Figura 2.26. Forma e dimensione, densità cromatica, colore, retinatura e orientamento.
Fonte: <http://civicingographics.ahref.eu>)

Le diverse tipologie di segni grafici possono essere utilizzate su *layer* differenti e in combinazione, in relazione a più variabili, dando vita a complesse mappe multivariate. È importante ricordare che molti dei dati presenti nel *geodatabase* non potranno essere visualizzati in mappa ma potranno essere analizzati e trattati comunque [*Ibidem*].

2.10. Quando condurre una GIS analysis

Raccogliere i dati più appropriati all'analisi GIS può assicurare al ricercatore una buona riuscita della stessa [Heywood *et al.*, 2006]. A tal fine, sarà necessario porsi alcune domande quali: «Esistono attributi riferiti alle entità spaziali identificate?»; «Se sì, quali sono?»; «Sono attributi riferiti allo spazio o

agli individui?»; «Potrebbero tali attributi influenzare l'intero processo di analisi GIS?»; «Esistono attributi che sono portatori di particolari significati per la comunità?».

Occorre rispondere a tali domande ancor prima di raccogliere i dati e di associare, ad ogni variabile, una specifica localizzazione geografica. Sarà indispensabile scegliere i confini che delimitano l'area geografica d'analisi. Questi potranno essere fisici (ad esempio, i confini amministrativi di una città) o concettuali (ad esempio, i luoghi ai quali le storie di una comunità si legano) [Goodchild, 2011].

A questo punto, il ricercatore potrà chiedersi: «Qual è la mia primaria unità di analisi?»; «Quali sono i confini fisici di tale unità?»; «Quali i suoi confini concettuali?»; «Qual è la natura di tali confini concettuali (economica, filosofica, sociale)?»; «Qual è il principale attributo geografico che verrà esaminato nel corso dell'analisi?»; «Tale attributo è collegato a una variabile dipendente o indipendente?».

2.11. Alcuni software GIS per l'analisi socio-spaziale

A metà anni Sessanta, l'Università di Harvard istituì il primo laboratorio di computer grafica e analisi spaziale. In quegli anni, i primi *software* per l'applicazione di metodi statistici territoriali, permettevano di gestire database, come il neonato *Census* americano, ma richiedevano ancora l'utilizzo di una grande quantità di memoria ad elevati costi di elaborazione [Clodoveu *et al.*, 2011]. Era il motivo per cui il gruppo di utilizzatori era composto principalmente da appartenenti ad organizzazioni universitarie e governative.

È solo a partire dagli anni Ottanta che tali applicazioni cominciarono ad essere utilizzate da utenti diversi per le analisi più disparate [ESRI, 2006]. Alla fine dei Novanta, grazie alla crescente disponibilità di *personal computer*, tra gli utenti GIS si annoverano già utenti specialistici, quali funzionari del governo federale, ricercatori universitari, analisti di mercato, pianificatori locali, ma anche utenti non esperti. Oggi, molte delle tecniche di analisi statistica vengono eseguite automaticamente dall'applicazione e gli utenti possono programmare da sé sub-

applicazioni GIS per soddisfare specifiche esigenze di analisi [Longley *et al.*, 2011].

Alcuni tra gli strumenti software più diffusi hanno reso più semplice il percorso di acquisizione di competenze analitiche socio-spaziali da parte degli scienziati sociali. Alcuni dei più comuni sono [Steinberg e Steinberg, 2006]:

- GeoDa™. Gratuito²⁰ e semplice da usare. È costituito da strumenti, dedicati all'analisi esplorativa multivariata e all'autocorrelazione spaziale locale e globale, utili a descrivere, mappare e analizzare i dati spaziali. Il pacchetto venne inizialmente sviluppato dal *Laboratorio di Analisi Spaziale* dell'Università di Illinois. Successivamente, si dedicarono allo sviluppo delle sue funzioni oggi più importanti i ricercatori del *GeoDa Center* dell'Università dell'Arizona [Anselin e Rey, 2006];
- STARS (*Space Time Analysis of Regional Systems*) è un pacchetto *open-source*²¹ progettato dalla *Free Software Foundation* (Boston), utile all'analisi dei trend temporali poiché gestisce i dati aggregati in unità areali per successivi periodi di tempo;
- CrimeStat® è un *software free*²² sviluppato da *Ned Levine & Associates* e finanziato dal *National Institute of Justice* americano. È un programma *Windows-based* e si interfaccia con numerosi software GIS allo scopo di fornire al ricercatore sociale strumenti di analisi statistica supplementari, utili ad analizzare e mappare i fenomeni criminali;
- ArcGIS è su software di proprietà della ESRI (*Environmental Systems Research Institute*). È uno dei più diffusi, dotato di funzionalità cartografiche per l'analisi spaziale. È possibile aggiungere moduli specialistici diversi con nuove funzionalità gestionali, integrandoli a quelli di base e utili, ad esempio, al calcolo, visualizzazione e automazione di processi complessi. Al momento della sua nascita, avvenuta negli anni Ottanta, era uno dei software più innovativi poiché la struttura del

²⁰ Liberamente scaricabile dal sito del GeoDa Center <http://geodacenter.asu.edu/>.

²¹ Liberamente scaricabile dal sito Sourceforge <http://sourceforge.net/projects/stars-py/>.

²² Liberamente scaricabile dal sito dell'ICPSR:
<http://www.icpsr.umich.edu/CrimeStat/download.html>.

geodatabase relazionale permetteva, sin da allora, di organizzare logicamente gli attributi e le informazioni geografiche connettendo oggetti spaziali e attributi qualitativi e quantitativi [Perencsik *et al.*, 2004];

- Quantum GIS è un'applicazione *open source*²³ sviluppata da un gruppo di volontari che si occupano di aggiornarla e correggerla. Rilasciato sotto la *GNU General Public License*, è un progetto ufficiale della *Open Source Geospatial Foundation* (OSGeo). Funziona su Linux, Unix, Mac OSX, Windows e Android e supporta numerosi formati vettoriali, *raster*, database e funzionalità. La sua interfaccia è tradotta in 14 lingue. Per arricchire le sue funzioni, è possibile far interagire col software dei *plugin*, caricabili all'apertura del programma [Steinberg e Steinberg, 2006].

²³ Liberamente scaricabile dal sito QGIS: <http://www.qgis.org/it/site/>

CAPITOLO TERZO

GEODATABASE

Un GIS nasce da un modello concettuale che permette di trasformare i fenomeni sociali in un costrutto rappresentativo, composto da entità spaziali portatrici di attributi. Prima che un modello implementato attraverso un software, è un modello mentale, elaborato tenendo conto anche dello schema logico che sta alla base del software stesso [Rigaux *et al.*, 2002]. Tale procedimento di trasformazione vincola la progettazione dell'indagine empirica a un controllo *ex ante* particolarmente strutturato che riduce il peso della sorte nel percorso della ricerca.

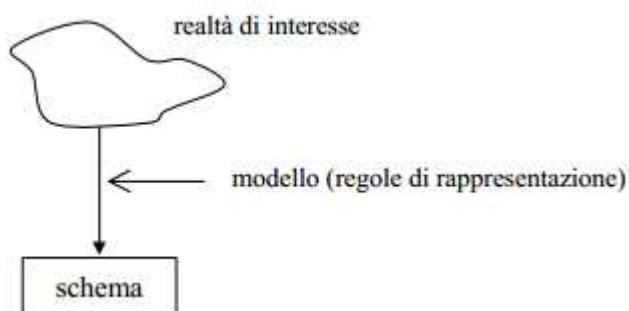
Il disegno dell'indagine e della progettazione GIS è esposto certamente a un maggior controllo e a una minore possibilità di fallimento poiché, affinché un GIS funzioni, deve essere necessariamente ben progettato logicamente [Hakim, 2000, 147] e orientato a criteri di razionalità strumentale. Scopriremo anche che, per un GIS *analist*, preparare il diagramma del modello concettuale è di fondamentale importanza per rispettare gli obiettivi della ricerca e delineare una strategia di indagine coerente. Rappresenterà la mappa di orientamento dalla quale tutti i ricercatori coinvolti potranno attingere e sulla quale iniziare a costruire il consenso sul modo di procedere [Creswell, 2003].

La GIS *analysis* per la ricerca sociale non è orientata ad un trattamento esclusivamente quantitativo dei dati. Piuttosto, integrandosi ad altre tecniche, ripristina uno *status* di pari dignità fra dati qualitativi e quantitativi, esponendo il processo della ricerca alla riflessività e all'apprendimento in corso. A tal fine, però, occorre primariamente che il ricercatore sia in grado di organizzare i dati nella struttura logica del geodatabase relazionale, conoscendo le regole che lo costituiscono e il linguaggio di interrogazione che permette la gestione dell'archivio dati [Schiavi, 2002].

3.1. Il modello dei dati

Assumendo che il ricercatore sociale abbia già stabilito che il GIS sia lo strumento più appropriato all'analisi del fenomeno di suo interesse, occorre che questi implementi il modello dei dati più adatto a rispettare il giusto equilibrio tra mondo reale da modellizzare e la sua astrazione [Clodoveu *et al.*, 2011].

Un modello dei dati, nell'ambito di una GIS *analysis*, si costruisce a partire dalle *entità spaziali* e individuando le *relazioni* che legano tra loro tali entità. Nel modello, il quale riduce la complessità del sistema, sono specificati tutti gli elementi che lo compongono, le relazioni fra le sue parti e le sue regole di funzionamento [Conte *et al.*, 2001].



(Figura 3.1. Modellizzazione della realtà)

In definitiva, l'insieme di *entità* e *relazioni* che descrive il fenomeno in esame costituisce il modello dei dati [Boffi, 2004, 2]. Prendersi del tempo da destinare alla sua elaborazione è indispensabile tanto quanto farlo per predisporre il *design* della ricerca. Anche in questo caso, è essenziale al fine di ridurre gli errori ed evitare le incongruenze.

La modellazione dei dati da assegnare alla GIS *analysis* passa attraverso una serie di operazioni logiche che è necessario eseguire con rigore [Steinberg e Steinberg, 2006, 140]:

1. Esame del fenomeno oggetto di studio;
2. Concettualizzazione della realtà che intendiamo modellizzare su *layer*;

3. Definizione delle relazioni logiche esistenti tra le componenti del modello;
4. Sviluppo, a partire dal modello concettuale e logico, del modello fisico *computer-based*.

L'incorporazione nel geodatabase di dati che non rispettano le regole del modello concepito dall'analista non condurrà a risultati della ricerca validi e attendibili poiché, piuttosto, veicoleranno tutt'altro significato rispetto agli obiettivi da raggiungere. I dati saranno appropriati all'analisi solo se il loro reperimento sarà stato ispirato dal modello concettuale il quale, a sua volta, permetterà di costruire un modello fisico funzionante [Schiavi, 2002].

Nei casi in cui l'analista sia in grado di identificare le entità spaziali più appropriate, e le relazioni che le contraddistinguono, ma non riesca a riprodurle in un *layer*, una delle soluzioni suggerite dagli autori è quella di testare il modello su un piccolo dataset e tentare di riconoscere eventuali cause di errori. Solo allora, dopo aver verificato la concreta efficienza del modello concettuale e logico, sarà possibile per lui addentrarsi nella costruzione fisica del geodatabase [Conte *et al.*, 2001].

3.2. Database relazionale

Il *database* è un archivio dati, o un insieme di archivi dati, che contiene informazioni strutturate e collegate tra loro in base a un particolare modello logico. L'organizzazione sistematica delle informazioni che lo costituiscono consente la gestione efficiente dei dati [Atzeni *et al.*, 2003]. Una base dati è, dunque, una raccolta di informazioni organizzata da un sistema di gestione database che risponde alle richieste dell'utente utilizzando le regole formalizzate in *query language* [Ullmann e Widom, 1997].

Inoltre, il database è una raccolta unificata e informatizzata di dati che può essere condivisa da più utenti, autorizzati ad utilizzare il software di gestione, e in grado di controllare, recuperare, manipolare, accedere ai dati al suo interno [Worboys, 1995].

Perché un sistema di archiviazione dati *computer-based* possa funzionare correttamente, restituendo informazioni rilevanti per il buon esito dell'indagine, deve essere affidabile. È importante che i dati vengano protetti da accessi non autorizzati, che siano al riparo da guasti imprevisti, che siano corretti, indipendenti e accessibili all'intero gruppo di ricerca [Dainelli *et. al.*, 2010, 41]. Caratteristica, quest'ultima, particolarmente rilevante per i sistemi di archiviazione di dati GIS utilizzati nel corso di una ricerca sociale. È indispensabile che gli utenti esperti, appartenenti ai diversi settori disciplinari, possano accedere al database interrogandolo per scopi diversi. Essi devono poter essere in grado di utilizzarlo, autonomamente e in gruppo, oltre che di recuperare e gestire informazioni in maniera ottimale [Conte *et al.*, 2001].

Il livello di prestazione di un sistema di gestione di una base dati aumenta o diminuisce in base alla quantità di dati incorporati e alla complessità con cui sono formulate le *query*. Quanto più è complessa l'interrogazione, tanto più è richiesto un livello di prestazione elevato²⁴. In particolare, è attraverso il database relazionale che è possibile avanzare le *query* più interessanti [Travaglini, 2004]. Si tratta di un modello di database la cui diffusione si fa risalire agli anni Settanta, ad opera di Edgar Codd, un ingegnere della IBM. La semplicità del linguaggio di interrogazione SQL (*Structured Query Language*)²⁵ [Dainelli, 2010, 46] fu decisiva per il suo successo. Attraverso il database relazionale, l'organizzazione fisica dei dati è affidata a modelli formalizzati. In questo modo, l'analista potrà preoccuparsi solo di progettare il modello concettuale, evitando di annotare informazioni riferite alla allocazione fisica dei dati.

Componente essenziale del database relazionale, le *tabelle*, chiamate anche *relazioni*, che possono contenere dati diversi, riferiti a individui, oggetti, eventi,

²⁴ Il database viene gestito dal DBMS (*DataBase Management System*) che funge da intermediario tra l'utente e i dati e svolge il lavoro di gestione fisica dei dati.

²⁵ La prima versione SQL venne elaborata a metà anni Settanta, sviluppata come strumento utile alla definizione dello schema della base dati e per l'inserimento, modifica e recupero dei dati del database. Nel 1986, divenne uno standard per i database relazionali. Viene definito linguaggio non procedurale poiché non richiede all'utente di definire le operazioni necessarie all'ottenimento del risultato ma di concentrarsi sugli obiettivi da raggiungere e, quindi, sugli scopi della domanda. È un linguaggio relazionalmente completo, nel senso che ogni espressione dell'algebra relazionale può essere tradotta in SQL.

ecc. Si tratta di vere e proprie matrici casi per variabili, che organizzano, collegano e aggregano i dati in modo strutturato e alle quali viene associata la colonna che localizza i dati²⁶. Per dare un'idea di come funzionino, riportiamo l'esempio dell'analisi sui beni confiscati alla mafia in Sicilia, condotta dal Dipartimento di Scienze Giuridiche, della Società e dello Sport dell'Università degli Studi di Palermo, nell'ambito del progetto "Mafia e impresa tra relazioni politiche e reti clientelari: alcuni casi studio nel territorio siciliano".

Le righe, definite anche *tuple* o *record*, rappresentano le unità di osservazione e consistono di una lista di valori, uno per ogni attributo. Gli attributi sono riportati nelle diverse colonne. L'ordine di posizionamento di righe e colonne non ha rilevanza e non altera il contenuto della relazione.

BENI							
Id Bene	Id mafioso	Region e Bene	Provincia Bene	Comune Bene	Data Confisca Definitiva	Ragione Sociale	Stato attività
TEST O/NUMERO	NUMERO	TESTO	TESTO	TESTO	DATA	TESTO	TESTO
62709	1	Sicilia	Palermo	Bagheria	02/04/87	DE.A	INATTIVA
206994	2	Sicilia	Palermo	Collesano	30/01/09	FATTORIA GARBINO GARA S.R.L.	INATTIVA

(Tabella 3.1. Tabella beni confiscati alla mafia in Sicilia)

Ogni cella contiene una sola occorrenza di uno specifico attributo, in modo da consentire l'identificazione univoca dei diversi *record*. I valori ammissibili per ogni attributo sono definiti da tutti quelli che questo può assumere; quelli contenuti, cioè, nel suo dominio [*Ibidem*, 2010].

²⁶ Spesso, le matrici mancano proprio del dato geografico. L'accorgimento di aggiungere la colonna degli indirizzi, o quella della provincia o della regione, rappresenta già una soluzione che può risolvere tale carenza. Sarebbe auspicabile che gli istituti di ricerca organizzassero i dati in forma di matrice considerando l'adiacenza e vicinanza delle unità di osservazione localizzabili geograficamente. Il *Census Bureau* americano, ad esempio, solitamente ordina le matrici in modo da accorpate insieme tutti gli stati della stessa regione. La classificazione interna alla regione segue, però, un ordine di tipo alfabetico. In questo caso, appare impossibile identificare le regioni in base alla loro prossimità geografica [Goodchild e Janelle, 2004, 3].

Ad esempio, l'attributo *Stato attività* ammette solo le stringhe alfanumeriche *Attiva solo per adempimento, Attiva sul mercato, In corso di aggiornamento, Inattiva, N.D.* Numeri interi, numeri reali, stringhe, date, ecc. sono alcuni dei domini dei valori che un certo attributo può assumere.

Non sempre i tipi di campi sono definibili chiaramente. Tra i più comuni, quelli che fanno riferimento al contenuto informativo, testuale, numerico, logico e geografico [Boffi, 2004, 4]. È l'analista a decidere attraverso quale tipologia di campo definire il valore da archiviare, in modo da gestirne adeguatamente il contenuto. Ad esempio, l'*Id Bene* è un codice identificativo che può essere gestito indifferentemente con campi testuale e numerico. La scelta genererà conseguenze diverse sui criteri di ordinamento e di ricerca.

Riportiamo la tabella proposta da Boffi [Ibidem, 5], che può essere utile consultare per scegliere tra i diversi campi disponibili nei database principali.

TIPO	SOTTOTIPO	CARATTERISTICHE
testo	testo	Testo, combinazioni di testo e numeri trattati come testo, fino a 255 caratteri
	memo	Come il campo testo, fino a 65535 caratteri
numerico	intero	Numeri interi compresi tra -32768 e 32767
	Intero lungo	Il range di validità è compreso tra +2 miliardi e -2 miliardi circa
	Precisione singola	Il range di validità è pari a numeri di circa 40 cifre con una precisione di 7 cifre
	Precisione doppia	Il range di validità è pari a numeri di circa 300 cifre con una precisione di 15 cifre
	decimale	Numeri con circa 30 cifre con una precisione elevata
	valuta	Una variante del formato decimale con precisione fino a numeri di 15 cifre nella parte intera e 4 nella parte decimale
calendario	data	Date di calendario, solitamente gregoriano, nell'arco di tempo di un secolo circa, per esempio dal 1930 al 2029
	ora	Ore con precisione fino al minuto secondo
logico	Si/no	Valori della logica booleana
Informazione multimediale	Immagini, disegni, video, musica, parlato	I database per uso individuale non contengono fisicamente tali oggetti ma collegamenti ad oggetti multimediali

(Tabella 3.2. Campi disponibili per la definizione dei domini degli attributi. Fonte: Boffi [2004])

Un database relazionale permette di generare un numero infinito di relazioni di diverso tipo. In particolare, esistono tre tipi di relazioni tra tabelle implementabili attraverso i software GIS [Ibidem]:

- Uno a molti;

- Molti a molti;
- Uno a uno.

La relazione uno a molti (1:N) è la più frequente in un database relazionale [Atzeni *et al.*, 2003]. Ad un *record* della tabella A possono corrispondere più *record* della tabella B. Mentre ad un *record* della tabella B corrisponde un solo *record* della tabella A.

Ad esempio, ad ogni comune elencato in tabella *Bene* possono fare capo più expropriari, unità di osservazione della tabella *Mafioso*.

Nella relazione molti a molti (N:M), ad un *record* della tabella A possono corrispondere più *record* della tabella B e, viceversa, ad un *record* della tabella B possono corrispondere più *record* della tabella A. La relazione genererà una nuova tabella, detta di congiunzione²⁷.

Nella relazione uno a uno (1:1), ad un *record* della tabella A può corrispondere un solo *record* della tabella B, e viceversa. Si tratta di relazioni poco usate, in quanto le informazioni delle due tabelle, facenti riferimento alle stesse unità di osservazione e contenenti classi di attributi diversi, possono essere contenute in un'unica tabella [Conte *et al.*, 2001].

Il database preso ad esempio potrebbe essere costituito da due diverse relazioni o tabelle: quella che definisce il bene (Tabella 3.1.) e quella che definisce l'ex-proprietario del bene (Tabella 3.3.). Le due tabelle sono caratterizzate dagli attributi il cui valore deriva dal dominio a cui è associato l'attributo stesso e la loro relazione è possibile grazie agli attributi in comune tra le due (ad esempio, *Id Bene*). La relazione *Bene-Mafioso* si traduce definendo in tabella un vincolo tra i dati relativi al bene (*Id bene, regione, forma giuridica, status azienda, ecc.*) e i dati relativi al mafioso (*nome, cognome, data di nascita, Id bene, ecc.*).

²⁷ Sulle relazioni multi-a-multi è da notare il fatto che non esistono efficienti strutture dati per la loro implementazione, spesso è richiesto di scomporre tali relazioni con varie relazioni multi-a-uno.

MAFIOSI			
Id Mafioso	Prevenuto	Data di nascita	Id Bene
1	GRECO MICHELE	12/05/1924	62709
2	BARATTA ANTONINO	25/08/1948	206994

(Tabella 3.3. Ex-Proprietari dei beni confiscati alla mafia in Sicilia)

3.3. Vincolo di integrità

Nella gestione del database relazionale, occorre sempre fare attenzione all'integrità dei dati. Essi devono essere affidabili, consistenti e accurati. Se il vincolo di integrità è soddisfatto, le risposte alle *query* avanzate al database riporteranno informazioni corrette [Dainelli, 2010, 50]. Esso può manifestarsi a livello di *tabella*, se il campo identificativo dei *record* è unico e sempre presente e, quindi, se ogni unità di osservazione è dotata di un identificativo che compare una sola volta nella stessa tabella. Si manifesta a livello di *campo*, se la struttura di ciascun campo è corretta e i valori riportati rispettano i vincoli di validità, in termini di accuratezza e consistenza. Si manifesta a livello di *relazione* (cosiddetta *integrità referenziale*), se la relazione tra le due tabelle è logicamente e fisicamente coerente. Infine, si manifesta a livello di *regole di funzionamento*, se impone un vincolo rispetto alla selezione dei dati o alla costruzione delle relazioni [Atzeni *et al.*, 2003]. Tra i vincoli di integrità più importanti che danno luogo a relazioni, la *chiave primaria*. Essa stabilisce che i valori non debbano essere ripetuti, in modo che non li si possa confondere con altri.

Considerata la tabella *Beni*, la chiave primaria potrebbe essere data dal sottoinsieme di attributi costituito da *Id Bene* e *Id Mafioso*²⁸. Nella tabella *Mafiosi*, invece, le chiavi primarie candidate potrebbero essere *Id Mafioso* e *Data di nascita*²⁹.

Un differente tipo di chiave è la *chiave esterna* che, al fine di costituire una relazione tra due o più tabelle, deve essere comune a tutte quelle coinvolte. Se decidiamo di generare tale relazione è perché riteniamo che le informazioni di una delle due tabelle diventerà completa solo facendo riferimento a quelle riportate nell'altra. Ad esempio, nelle due tabelle *Beni* e *Mafiosi* ricorrono due attributi: *Id Bene* e *Id Mafioso*, due identificativi composti da numeri interi. La presenza di valori omologhi in tabelle differenti rende possibile il riferimento.

3.4. Join, proiezione e selezione

Le chiavi servono a gestire, manipolare i dati e svolgere particolari operazioni, quali la combinazione o *join*. Date due tabelle caratterizzate da alcuni specifici attributi, la nuova relazione sarà costituita dalla combinazione tra tutte le tuple che ricorrono in entrambe le relazioni.

La *proiezione* si applica su una singola tabella o relazione e permette di estrarre da essa un sottoinsieme di attributi [Kaushik *et al.*, 2004]. Ad esempio, potremmo presentare al database un'istanza di proiezione sulla relazione *Bene* chiedendo che ci restituisca l'*Id Bene* e la *Ragione Sociale*. Passando in rassegna gli attributi interessati, la procedura restituirà i valori registrati nelle tuple corrispondenti.

L'operazione di *selezione* (o *restrict*), invece, agisce primariamente sulle tuple [Conte *et al.*, 2001]. Considerata la tabella *Mafiosi*, potremmo presentare istanza di selezione chiedendo al database di restituirci tutti i valori uguali a "*michele greco*", in modo da verificare quali siano i beni confiscati al mafioso.

²⁸ La scelta ricade su due attributi poiché, nel caso in cui il mafioso abbia posseduto più beni, una sola chiave potrebbe non essere sufficiente a garantire l'univocità del riconoscimento.

²⁹ Si ricorre ad entrambi gli attributi poiché è possibile che la data di nascita di due mafiosi sia la stessa.

La selezione può prendere in considerazione più di una condizione attraverso l'utilizzo degli operatori logici AND, OR, NOT e selezionare, ad esempio, i *record* riguardanti i beni appartenenti a "michele greco" che si trovano in stato INATTIVO.

3.5. Operazioni sui dati con SQL

I GIS sono dotati di interfacce *friendly* che permettono di interrogare i dati in maniera facile e intuitiva. Tuttavia, nonostante non sia necessario per il ricercatore sociale approfondire le competenze sull'utilizzo della sintassi SQL, daremo qui qualche suggerimento utile alla comprensione della struttura logica delle *query*, rimandando ad altra bibliografia per ulteriori approfondimenti [Belussi, 2006].

Le operazioni sui dati si traducono in interrogazioni all'archivio. Le risposte alle istanze (o *query*), formulate dall'analista, restituiranno nuove variabili, unità e relazioni, attraverso operazioni logiche [Ozsu, 1994]. La struttura base dell'interrogazione è costituita da un verbo, che definisce l'operazione da seguire, dal riferimento ai dati su cui operare, dagli obiettivi a cui l'interrogazione stessa è finalizzata. Tra gli elementi che compongono una *query*, sono di fondamentale importanza le parole chiave che specificano il contenuto dell'interrogazione [Intesa GIS, 2004].

Si consideri il comando generico di recupero dati *SELECT FROM* che restituisce una lista di campi appartenenti a un certo numero di tabelle.

```
SELECT variabile 1, variabile 2, variabile n  
FROM database
```

Per capire tale struttura sintattica, basta ricordare l'operazione di proiezione di cui abbiamo discusso in precedenza. Si consideri la tabella *Beni*. Formulando la *query*:

```
SELECT
```

Id Bene, Stato Attività

FROM

Beni

il verbo *SELECT* effettua una proiezione degli attributi *Id Bene* e *Stato Attività*. *FROM* indica la tabella dalla quale verranno estratti gli attributi (*Beni*). Il risultato sarà dato da una nuova tabella riportante esclusivamente gli attributi derivanti dalla proiezione.

Con il comando *SELECT*, si possono effettuare anche delle operazioni di *join*.

Ad esempio, nella *query*:

SELECT

Beni, Id Bene,

Beni, Stato attività,

Mafiosi, Id Mafioso,

Mafiosi, Prevenuto

FROM

Beni, Mafiosi (la tabella venuta fuori da un'eventuale operazione di *join*)

WHERE

Beni, Id bene = Mafiosi, Id Mafioso

la proposizione *WHERE* restituisce il *join* poiché specifica che i *record* delle due relazioni *Beni* e *Mafiosi* debbano essere combinate solo se i valori degli attributi *Id Bene* e *Id Mafioso* sono uguali. Correlando le informazioni presenti nelle due tabelle, sarà possibile capire chi è il mafioso e quale bene gli sia stato confiscato. Tramite le *query*, è possibile inoltre aggregare le unità di osservazione. Ad esempio, se consideriamo la tabella *join* (*Beni, Mafiosi*), l'aggregazione di soggetti che possedevano un bene localizzato in uno dei comuni della provincia di Palermo si realizza specificando, in una *query*, una clausola di raggruppamento:

SELECT

Join. Prevenuto

FROM

Join

GROUPED BY

Join. Provincia Bene.

3.6. Modello georelazionale

In un GIS, l'informazione alfanumerica è sempre associata a quella geometrica. Il database GIS, infatti, è un DBMS potenziato poiché possiede anche capacità di gestione dei dati spaziali [Güting, 1994]. È dotato, inoltre, di un linguaggio di interrogazione che possa agire sui punti, le linee e i poligoni, componenti base per la modellazione delle strutture e relazioni spaziali.

Il modello di dati di riferimento a livello europeo per i database spaziali è il *Simple Feature Access*, concepito dal *Consorzio OpenGIS®*. Tale standard prevede che gli insiemi di elementi spaziali vengano memorizzati nel geodatabase come tabelle di database relazionale e che ogni oggetto spaziale sia registrato come *record* della tabella [*Open Geospatial Consortium*³⁰, 2010]. Nel caso del modello georelazionale, la tabella assume il nome *feature table*.

Mentre gli attributi non spaziali saranno memorizzati utilizzando le regole del linguaggio SQL [ESRI, 2006], quelli spaziali saranno memorizzati in base al modello geometrico di riferimento. In questo modo, gli operatori relazionali spaziali estendono le capacità di interrogazione dei database tradizionali, manipolando anche i dati geografici. I principali operatori nella GIS *analysis* sono [Dainelli, 2010, 70]:

- *Equals*: individua l'uguaglianza tra due oggetti;
- *Disjoint*: verifica se due oggetti sono disgiunti e se, quindi, non hanno punti in comune;
- *Touches*: individua gli oggetti che si toccano in punti esterni;

³⁰ L'OGC è un'organizzazione internazionale che cerca di definire e dettare le linee guida sui formati e sugli standard di accesso ai dati, nel tentativo di arrivare a un'interoperabilità sempre maggiore nella condivisione dell'informazione territoriale.

- *Within/Contains*: verifica che un oggetto sia interamente contenuto da un altro oggetto;
- *Overlaps*: verifica che i due oggetti siano sovrapposti.

	Poligono/Poligono	Linea/Linea	Linea/Poligono	Linea/Punto	Punto/Poligono	Punto/Punto
Disjoint						
Touches						
Within/Contains						
Overlaps						

(Figura 3.2. Combinazione tra operatori e geometrie. Fonte: Dainelli, 2010, 70)

Quello presentato non è un elenco esaustivo ma è un buon punto di partenza per cominciare a formalizzare, nel modello logico, le interazioni tra le entità nello spazio [Dainelli, 2010].

3.7. Costruzione del modello concettuale e logico *benidiscilia*

Per lo svolgimento dell'esercitazione proposta, si è scelto di utilizzare il software ArcGIS 10, disponibile in versione di prova per 60 giorni, se richiesto a ESRI Italia³¹.

Come già chiarito, il primo passo da compiere nella costruzione del geodatabase è la definizione dello schema concettuale e logico [Date, 2003]. Occorrerà decidere

³¹ www.esriitalia.it

cosa rappresentare nel database, quali funzioni questo dovrà svolgere e quali rapporti dovranno esistere tra le entità registrabili. È importante ribadire che tali scelte dovranno essere organizzate secondo un meta-linguaggio formale, in modo che sia universalmente comprensibile e privo di ambiguità anche per chi vorrà utilizzarlo nel corso di indagini successive. Qui, si utilizzerà una versione semplificata di uno dei linguaggi più diffusi nella costruzione dei modelli concettuali: l'UML (*Universal Modeling Language*)³² [Thomas, 2003].

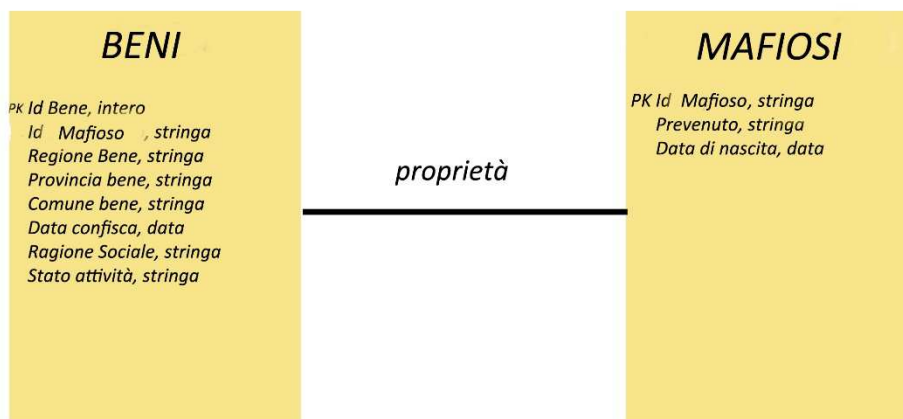
L'UML è composto da diversi diagrammi, o viste, che rappresentano il sistema da geocodificare. Il *diagramma delle classi* aiuta a definire classi di oggetti caratterizzate da attributi e relazioni tra attributi [Fowler, 2004].

Nel linguaggio UML, una classe viene rappresentata da un rettangolo composto da tre sezioni: la prima, in alto, riporta il nome della classe; al centro vengono elencati gli attributi della classe e i rispettivi domini (intero, stringa, ecc.). Almeno uno degli attributi dovrà fungere da identificatore univoco e dovrà essere contraddistinto dalla sigla PK (*Primary Key*) [Brodeur *et al.*, 2000]; all'ultima sezione, può essere dedicato lo spazio che si trova immediatamente sotto l'elenco degli attributi, nel caso in cui sia necessario descrivere il comportamento della classe. Fra le classi, si instaurano rapporti di associazione rappresentati da una linea retta che le unisce.

Considerata la piccola base dati ottenuta dall'Agenzia dei Beni Confiscati³³, si mostrerà come implementare un semplice schema concettuale di un database contenente, tra le entità, anche quelle geolocalizzabili. Il sistema conterrà, come discusso, informazioni sulle aziende confiscate (*nome, ubicazione, destinazione d'uso, stato attività, ecc.*) e informazioni sui mafiosi ex-proprietari, (*prevenuto, data di nascita, ecc.*). La Figura 3.3. descrive la relazione esistente tra il bene e il mafioso ex-proprietario, in UML.

³² Nacque intorno alla metà degli anni Novanta. Non rappresenta solo uno strumento utile alla progettazione di database ma viene anche utilizzato nell'organizzazione di progetti. La versione che rappresenta la realtà dei dati geografici è la Geo-UML, realizzata dal consorzio italiano Intesa GIS. Tuttavia, qui si ritiene di non dover approfondire le modalità di utilizzo di tale linguaggio formalizzato poiché supera gli intenti del presente lavoro.

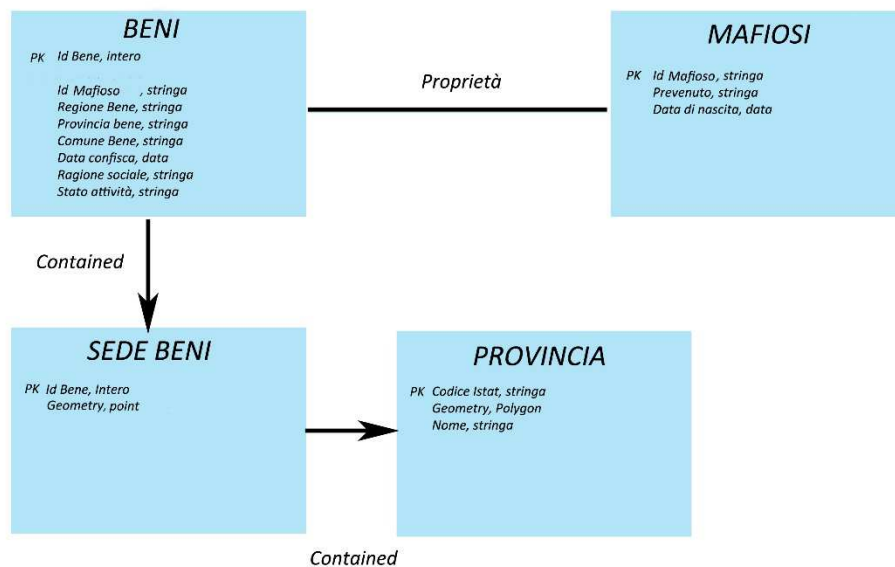
³³ <http://www.benisequestraticonfiscati.it>



(Figura 3.3. Relazione tra classi in UML)

Una classe può rappresentare oggetti geometrici se possiede almeno un attributo di tipo geometrico. Nel presente caso, dato che lo scopo è quello di costruire un database che contenga al suo interno informazioni sui beni localizzati e su chi ne deteneva la proprietà prima della confisca, si potrà decidere di inquadrare le ubicazioni dei beni (punti) in un contesto amministrativo provinciale (poligoni). Pertanto, la rappresentazione spaziale passa dalla definizione, stavolta, di quattro classi di oggetti:

- *Beni*, contenente attributi non spaziali;
- *Sede beni*, in cui è necessario specificare un attributo geometrico che contenga valori di tipo puntuale, in quanto il bene è localizzabile attraverso il suo indirizzo;
- *Mafiosi*, che è di tipo non spaziale;
- *Provincia*, che rappresenta una seconda classe di oggetti geometrici di tipo superficie, rappresentante le provincie. Fra *Beni* e *Provincia* esiste una relazione di contenimento, poiché i beni sono contenuti all'interno del territorio provinciale. La relazione scelta è di tipo *Contain*.



(Figura 3.4. Modello concettuale con UML semplificato)

Tale modello concettuale dovrà essere, poi, trasformato in modello logico contenente lo schema del database e le relative tabelle o relazioni. Questa trasformazione avviene scegliendo un appropriato modello dei dati. Qui, si è scelto di operare attraverso il modello dati relazionale, di cui si è già discusso. Verranno applicate, perciò, alcune regole di derivazione per le classi e per le associazioni, valide all'interno di tale modello [Dainelli, 2010, 83]:

- Ogni classe di oggetti diventa una tabella, o relazione;
- Ogni attributo di una classe di oggetti diventa un attributo della relazione;
- Ogni colonna della tabella eredita le caratteristiche dell'attributo della classe di oggetti da cui deriva;
- L'identificatore PK della classe diventa la chiave primaria della tabella derivata.

Consideriamo le due classi di oggetti *Beni* e *Mafiosi*. Gli attributi *Id Bene* e *Id Mafioso* sono i corrispondenti identificatori univoci. L'associazione tra le due entità comporta la presenza, nella tabella *Bene*, dell'attributo *Id mafioso*, chiave esterna che garantisce il legame. La relazione spaziale *Contained* non sarà esplicitata nel modello logico ma sarà implementata grazie alla funzione del

software. In particolare, in ArcCatalog, applicazione di ArcGIS 10, gli operatori relazionali sono applicabili con la funzione di *query* spaziale *Select by location*.

La maggior parte dei software GIS permette, inoltre, di organizzare, nel modello fisico, la struttura definitiva del modello logico inserendovi [Brodeur *et al.*, 2000]:

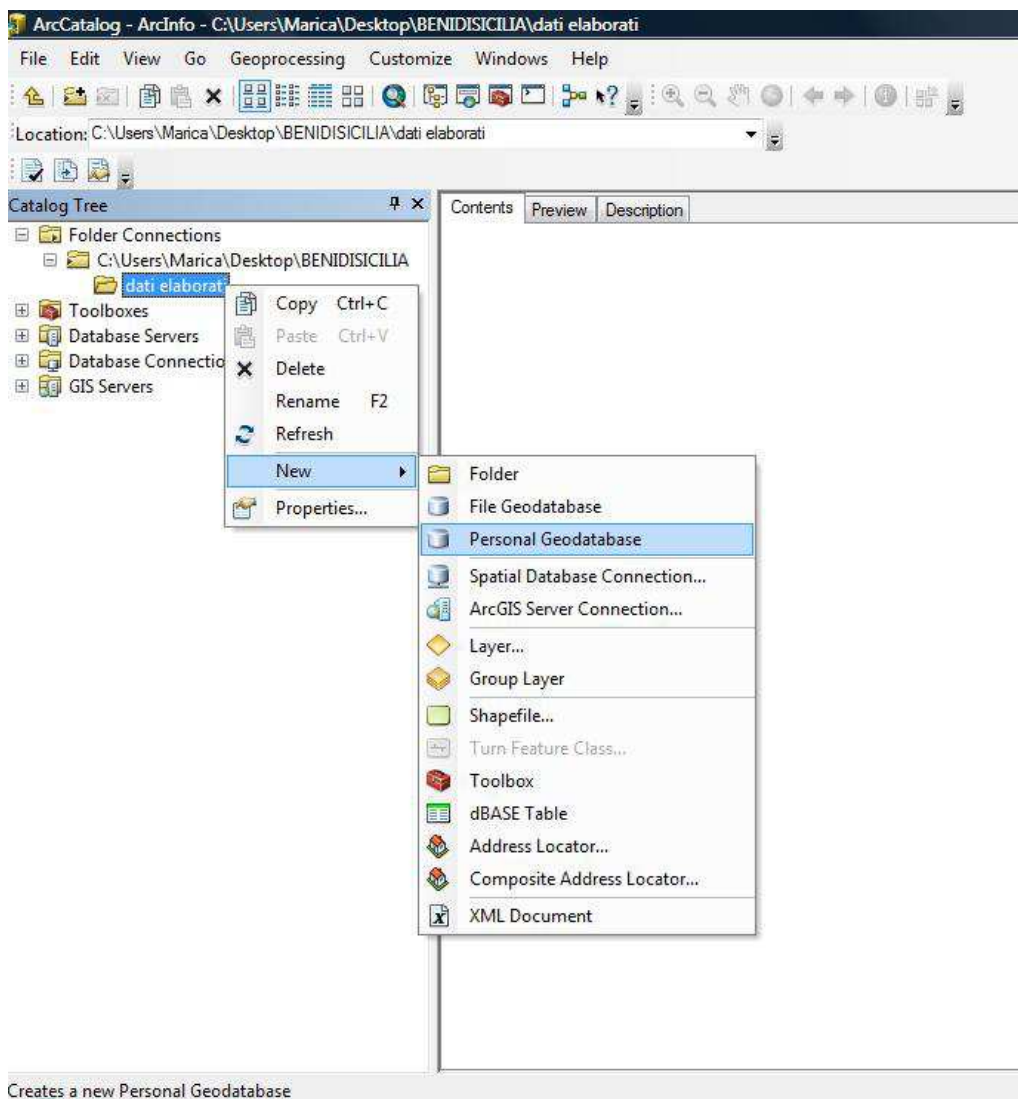
- *layer* di primitive geometriche puntuali, lineari e poligonali (*Feature Class*) costituiti da elementi appartenenti ad un singolo genere di primitiva;
- la tabella degli attributi relativa.

Le *Feature Class*, insomma, sono organizzate in insiemi di *Feature Class*, denominati *Feature Dataset* e tabelle (*Table*).

Nel caso dell'indagine Beni di Sicilia, le tabelle SEDI BENE e PROVINCIA rappresentano oggetti geometrici: nel geodatabase saranno, quindi, tradotte in *Feature Class*. Le relazioni BENI ed MAFIOSI saranno, invece, rappresentate da oggetti di tipo *Table*. Tutte le *Feature Class* saranno organizzate all'interno di un singolo *Feature Dataset*.

Si crei una cartella e la si denomini BENIDISICILIA. Al suo interno, si crei un'altra cartella denominata DATI_ELABORATI. Si apra l'applicazione ArcCatalog di ArcGIS 10. Si clicchi, col tasto destro del mouse, sulla cartella presente nel *Catalog Tree* denominata *Folder Connections* e si scelga la cartella BENIDISICILIA. Si clicchi con il tasto destro del mouse sulla cartella DATI_ELABORATI contenuta nella cartella BENIDISICILIA, e dal menu contestuale, Si scelga *New / Personal Geodatabase*. Lo si denomini subito *Gdb_benidisia.mdb*³⁴.

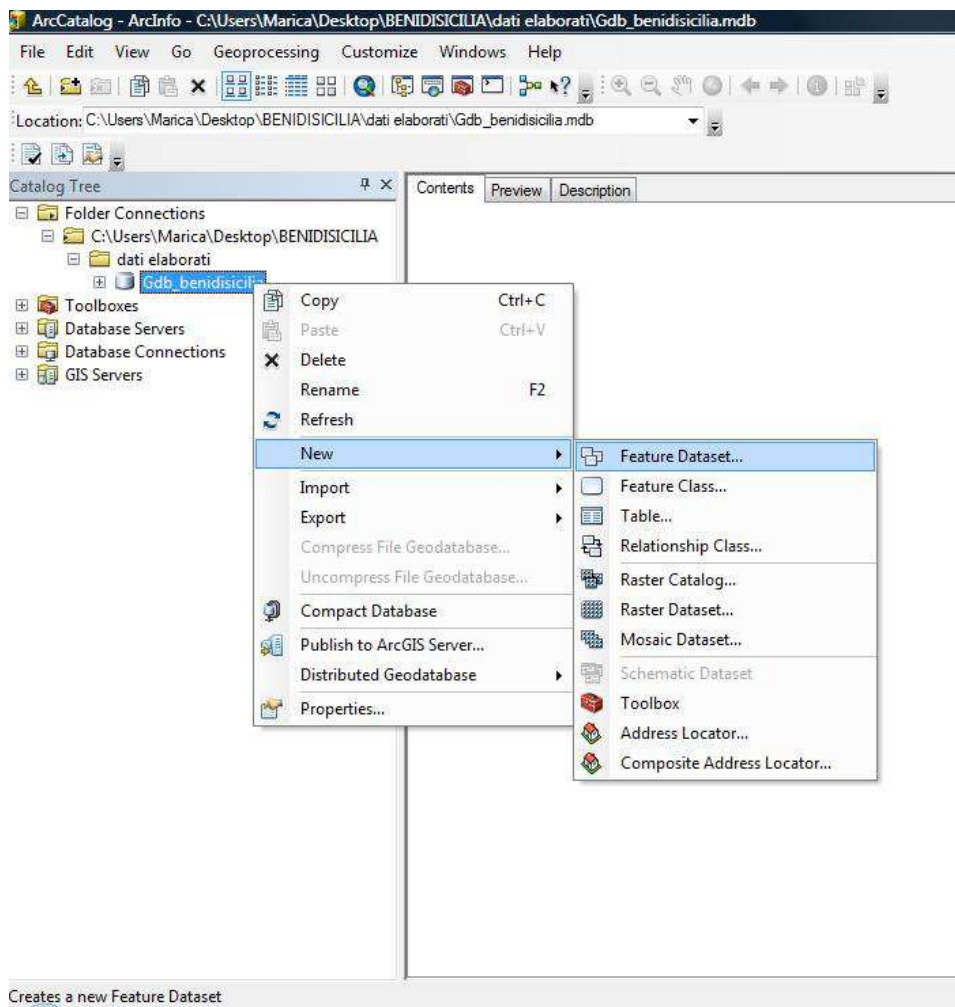
³⁴ L'estensione .mdb è la tipologia usata anche da Microsoft Access.



(Figura 3.5. Creazione in ArcCatalog del personal geodatabase)

Il Gdb (geodatabase) è ancora un oggetto vuoto; occorrerà inserire al suo interno le componenti quali *Feature Dataset*, *Feature Class* e le tabelle, in base allo schema logico descritto in precedenza.

Per definire il *Feature Dataset*, si impostino i parametri relativi al sistema di riferimento. Facendo clic col pulsante destro del mouse su *Gdb_benidiscilia.mdb*, si selezioni il comando *New / Feature Dataset*.

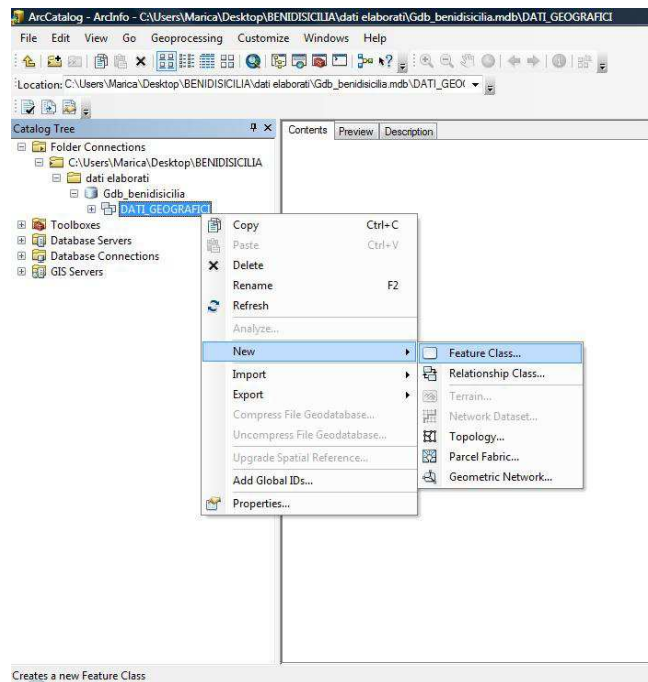


(Figura 3.6. Creazione nel geodatabase di un nuovo Feature Dataset)

Si nomini il nuovo *Feature Dataset* DATI_GEOGRAFICI.

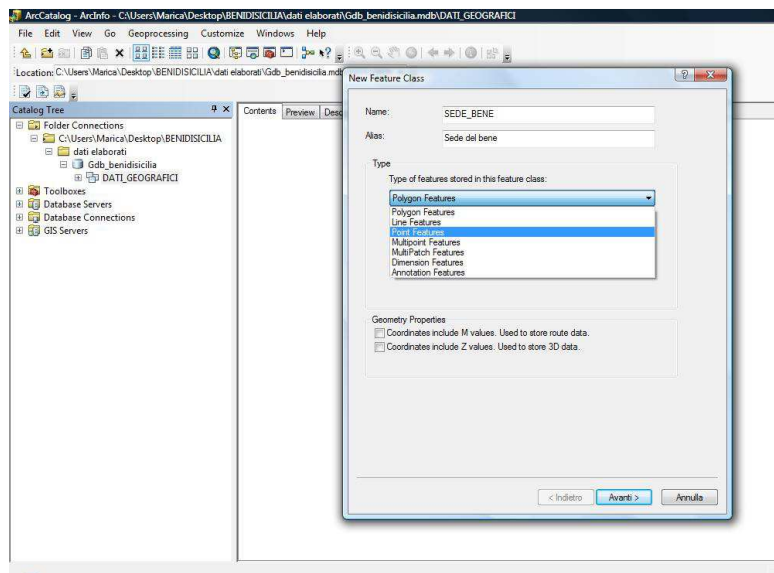
Nella finestra SPATIAL REFERENCE, occorrerà impostare il sistema di riferimento. Perciò, da *Projected Coordinate Systems*, si selezioni *National Grid* e si scelga *Monte Mario Italy 1*. Si clicchi su OK per applicare le impostazioni e si chiuda la finestra.

Dopo le *Feature Dataset*, si creino le *Feature Class*. Si clicchi, perciò, con il tasto destro del mouse sul *Feature Dataset* DATI_GEOGRAFICI e si scelga *New / Feature Class*.



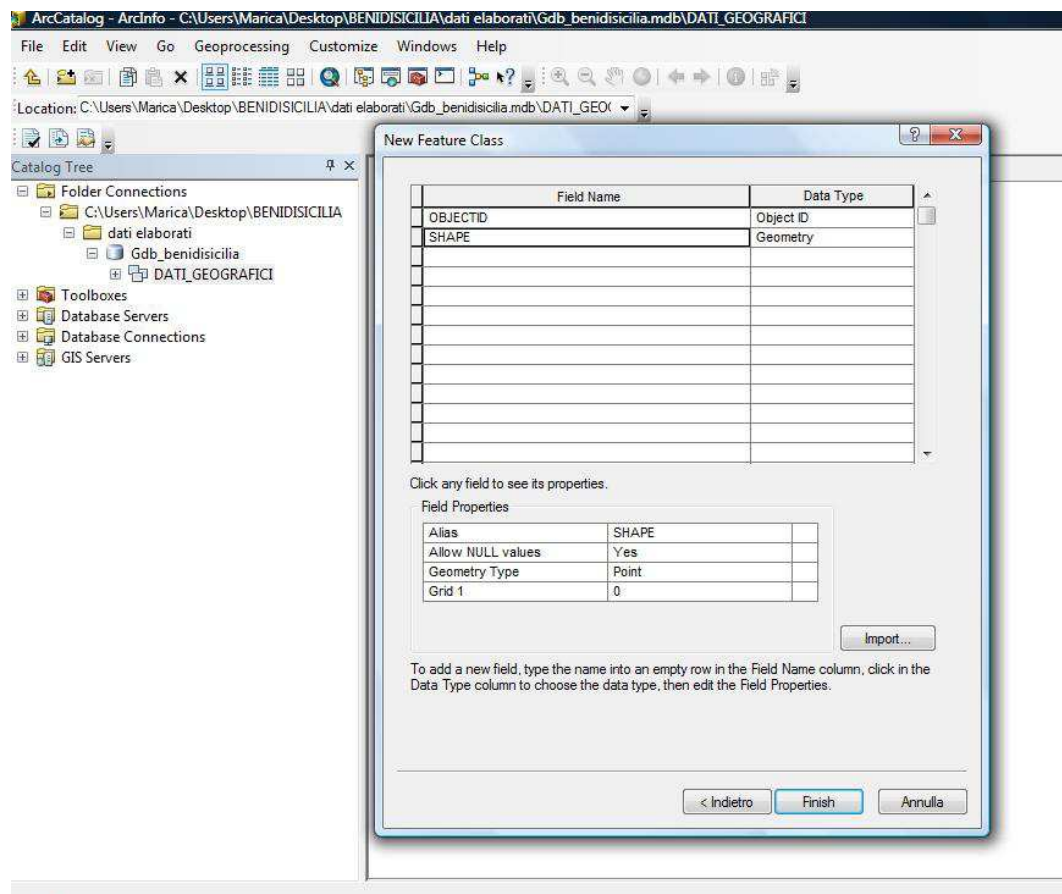
(Figura 3.7. Creazione nel feature dataset di una nuova Feature Class)

Il *wizard* guiderà l'analista nelle diverse fasi della creazione dell'oggetto. La prima *Feature Class* che si intende creare è SEDE_BENE.



(Figura 3.8. Creazione feature class SEDE_BENE)

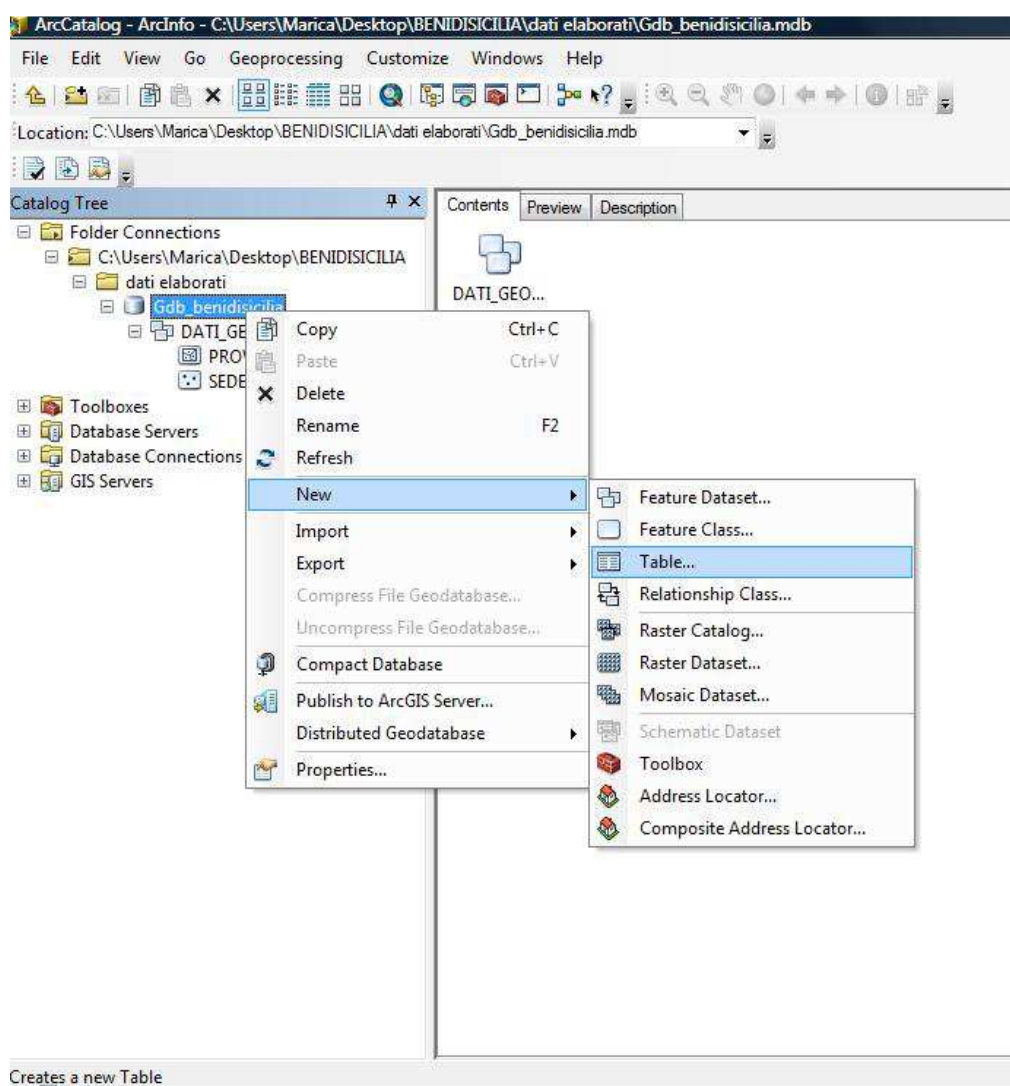
Si impostino, perciò, il nome e l'Alias "*Sede del bene*". Si scelga il tipo di dato puntuale, dato che si tratta di una classe riportante coordinate. Cliccando su AVANTI, si aprirà la finestra di definizione degli attributi. ArcGIS impone di *default* due attributi: OBJECTID, che è un codice identificativo numerico intero positivo univoco, e SHAPE, che definisce la geometria della classe. Se si clicca su SHAPE, in *Field Properties* verranno mostrate le proprietà dell'attributo. In particolare, *Geometry Type* descrive la primitiva geometrica (in questo caso *Point*) che sarà utilizzata per memorizzare gli oggetti della classe.



(Figura 3.9. Data Type e Properties della feature class)

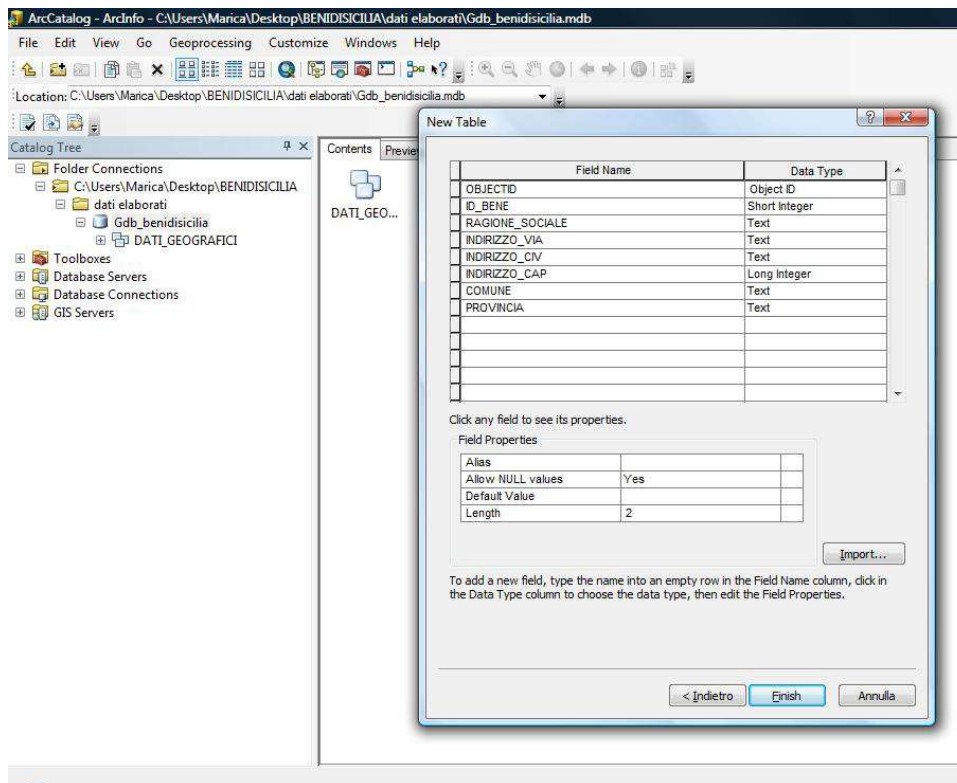
Gli altri attributi della classe possono essere memorizzati cliccando sulle righe vuote, scrivendo il nome in *Field Name* e il tipo di dato nella colonna immediatamente a destra. Nel caso dell'attributo SEDE_BENE, occorre aggiungere il campo ID_BENE e indicare come tipo *Short Integer*, dato che nella

costruzione del modello logico si è stabilito che il valore debba essere rappresentato da un numero intero. Si clicchi su FINE e si prosegua con la definizione della seconda *Feature Class* denominata PROVINCIA. In questo caso l'attributo SHAPE è un poligono. Si aggiunga agli attributi l'ID PROVINCIA (*short integer*). Si clicchi su FINE e si passi alla definizione delle tabelle. Si clicchi con il tasto destro del mouse su *Gdb_benidiscilia.mdb*. Dal menu contestuale, si selezioni *New / Table*.

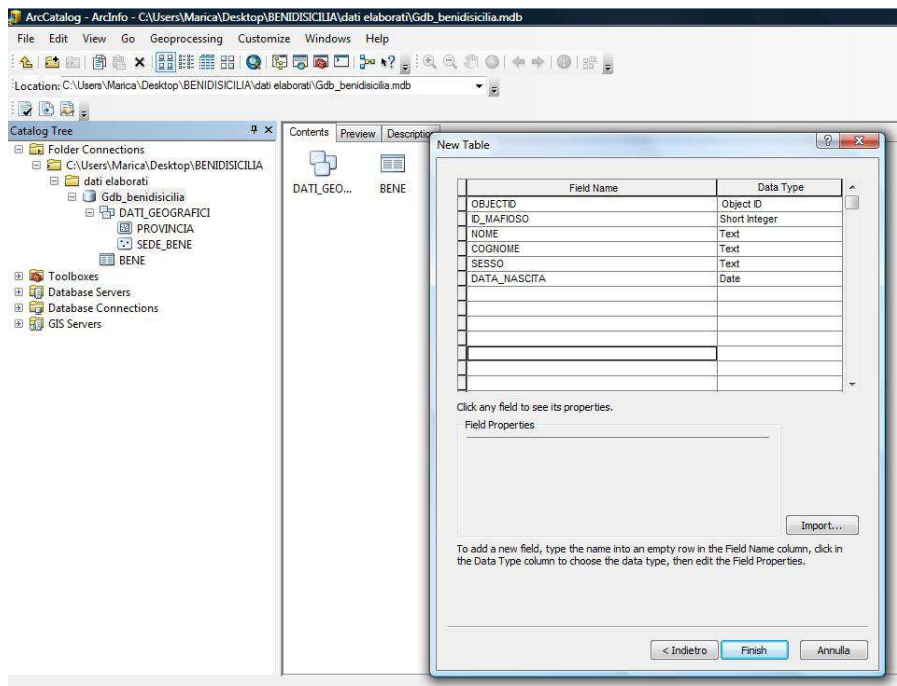


(Figura 3.10. Creazione nel geodatabase di una nuova tabella)

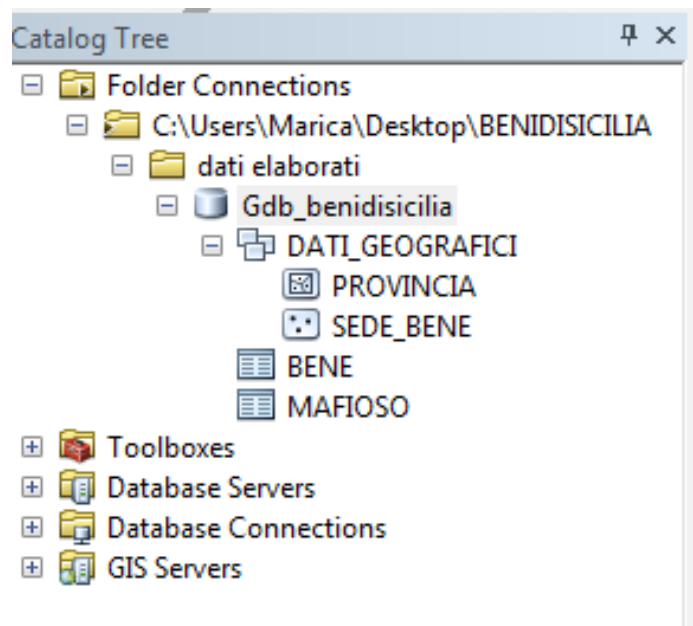
Anche qui, si definiscano nome, *Alias* della tabella e relativi attributi.



(Figura 3.11. Data type e properties della tabella Bene)



(Figura 3.12. Data type e properties della tabella Mafioso)

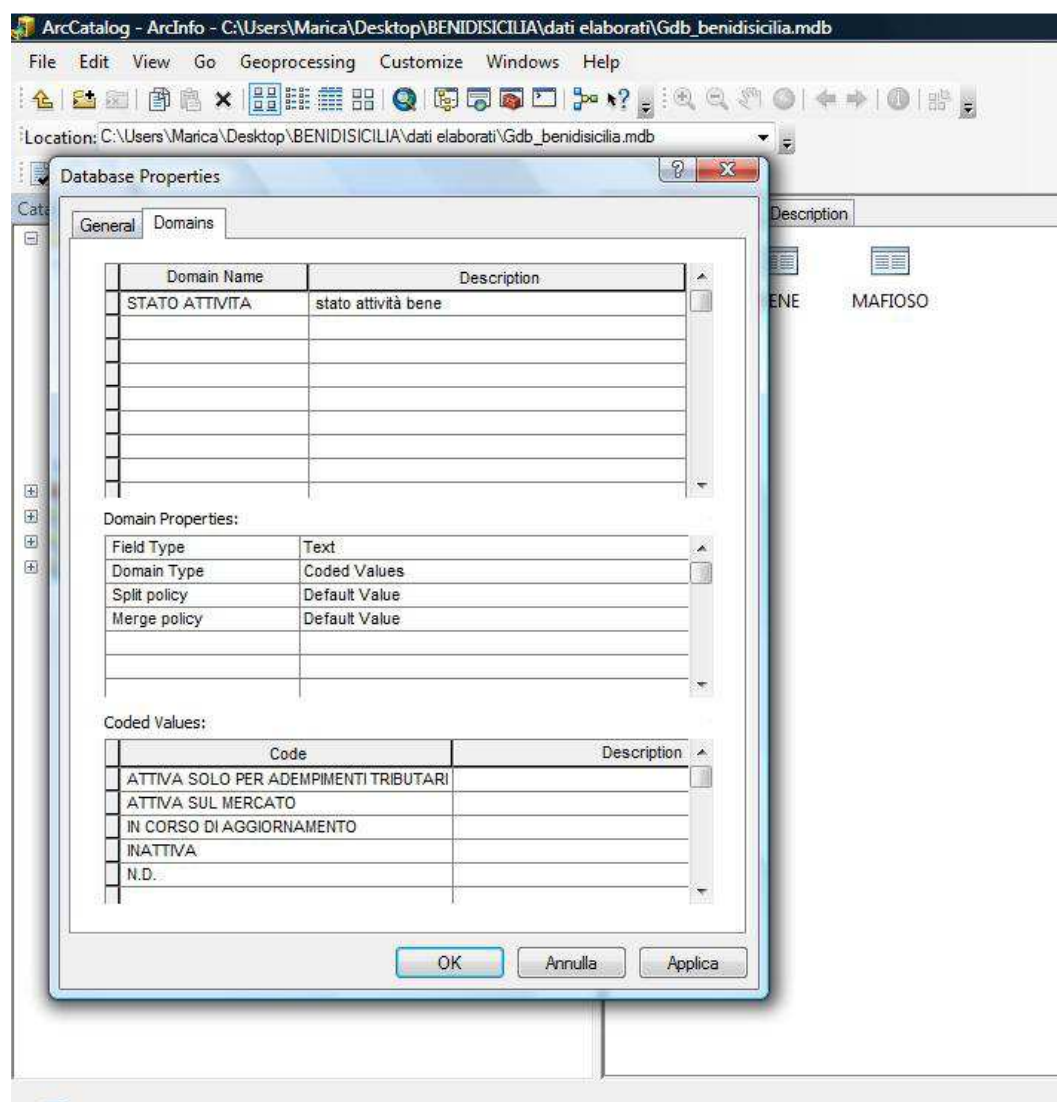


(Figura 3.13. Struttura complessiva del geodatabase)

È arrivato il momento di chiarire come avviene il processo di definizione dei domini degli attributi, attraverso il quale i valori codificati, o gli intervalli di valori ammissibili, possano essere associati agli attributi stessi. In questo modo, nel processo di popolamento della tabella o della classe, sarà sufficiente scegliere tra i valori associati in precedenza, piuttosto che digitarli di volta in volta.

Si definisca il dominio dell'attributo STATO ATTIVITA appartenente alla tabella BENI. Si clicchi, con il tasto destro del mouse, su *Gdb_benidiscilia.mdb* e si selezioni *PROPERTIES*. Dalla finestra *Database Properties*, si selezioni la scheda *DOMAINS*. La parte superiore della scheda accoglie i nomi dei domini che si intendono definire. Nella prima cella della sezione *Domain Name*, si digiti *STATO ATTIVITA*. Nella parte centrale della scheda, vengono specificate le proprietà del dominio. Si scelgano, in *field type* e *domain type*, rispettivamente *Text* e *Coded Values*. Nella parte bassa della finestra, vanno inseriti i valori codificati. Si digiti, nella colonna *Code* i valori *ATTIVA SOLO PER ADEMPIMENTI TRIBUTARI*, *ATTIVA SUL MERCATO*, *IN CORSO DI AGGIORNAMENTO*, *INATTIVA*, *N.D.*

Si clicchi su APPLICA. È possibile ripetere l'operazione per la definizione di tutti i parametri dei domini di interesse dell'analista.

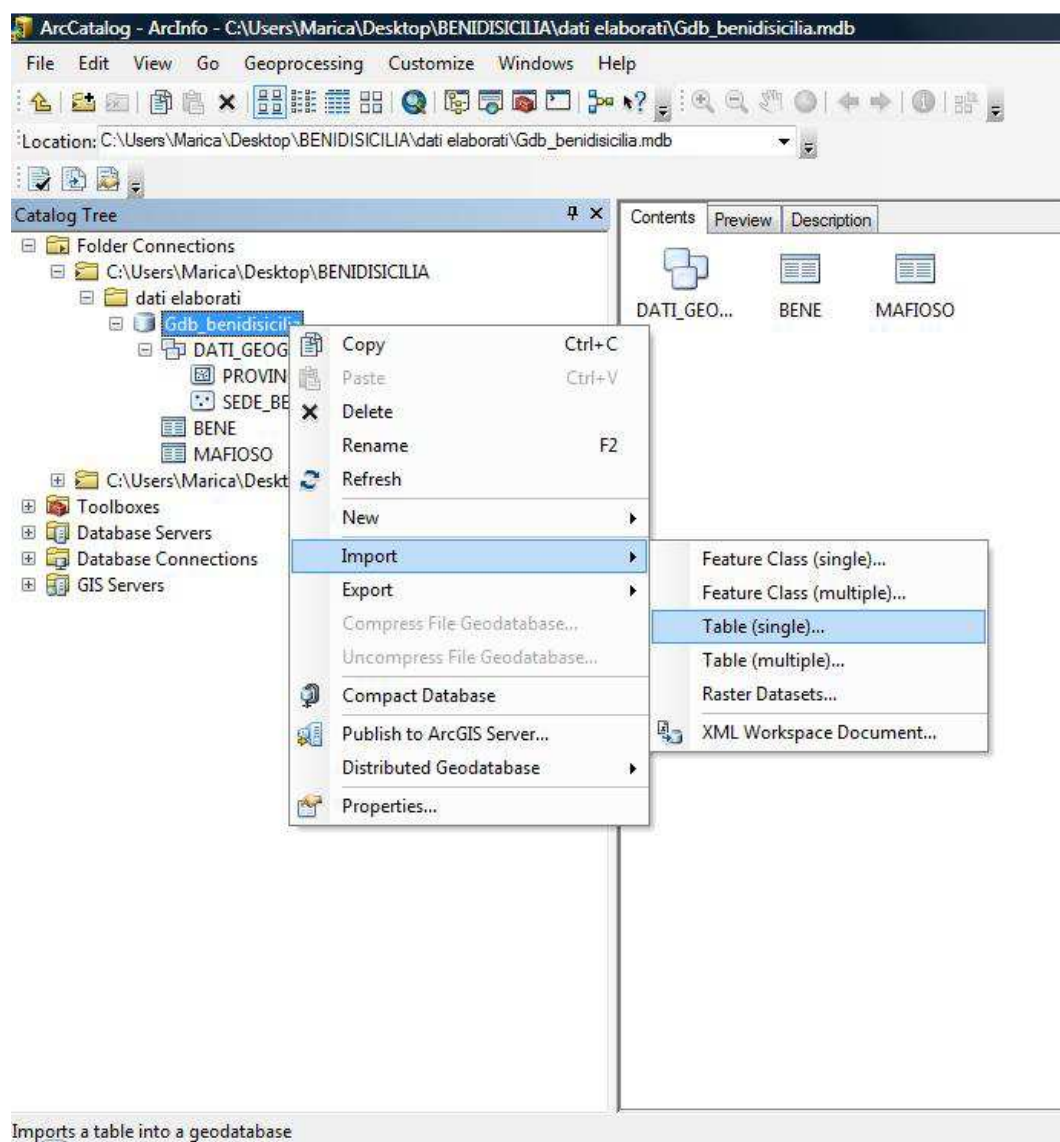


(Figura 3.14. Definizione dei domini nel geodatabase)

Qui, si è scelto di tralasciare i dettagli sulle operazioni di popolamento del geodatabase nel corso delle quali occorre digitalizzare le *Feature Class*, a partire da cartografia esistente o da dati rilevati *in situ* (ad esempio, da rilevamenti GPS), e le tabelle, a partire dagli attributi nelle celle³⁵. Si ritiene, piuttosto, di dover

³⁵ Rimandiamo, per ulteriori approfondimenti, ai manuali dedicati ai GIS presenti in bibliografia e alla guida ArcGIS.

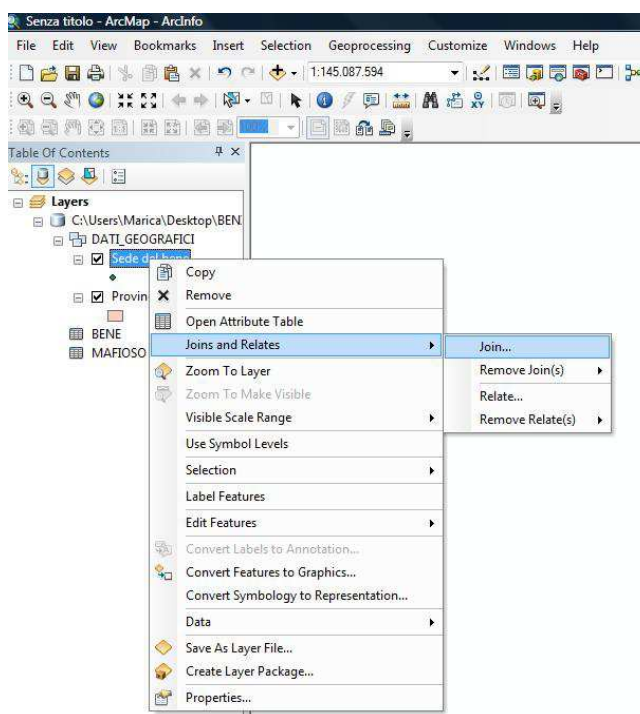
sottolineare che è possibile importare dati nello schema utilizzando il comando LOAD DATA richiamabile cliccando, col tasto destro del mouse, sulla *Feature Class* o Tabella da popolare. In questo caso, i dati prenderanno la forma dello schema precostituito. Cliccando, con il tasto destro del mouse, sul geodatabase sarà possibile, invece, scegliere il comando IMPORT che permette di importare il dato nella forma in cui si trova e senza possibilità di adattarlo alla struttura del geodatabase stesso.



(Figura 3.15. Importazione dati nel geodatabase)

Dopo aver caricato i dati cartografici nelle *Feature Class* (nel caso dell'indagine Beni di Sicilia, file vettoriali contenenti le coordinate che localizzano i beni³⁶) e nelle tabelle, si mostrerà come analizzare le relazioni esistenti tra i dati attraverso la funzione *Join* dell'applicazione ArcMap di ArcGIS 10.

Si carichi in ArcMap il *Gdb_benidiscilia.mdb* e si aggiungano, attraverso il pulsante *Add Data*, tutte le tabelle e le *Feature Class* presenti nel geodatabase. Per effettuare il *Join*³⁷ tra la *Feature Class* SEDE BENI e la tabella BENI, si clicchi, con il tasto destro del mouse, su SEDI BENE e si scelga il comando *Joins and Relates / Join*.

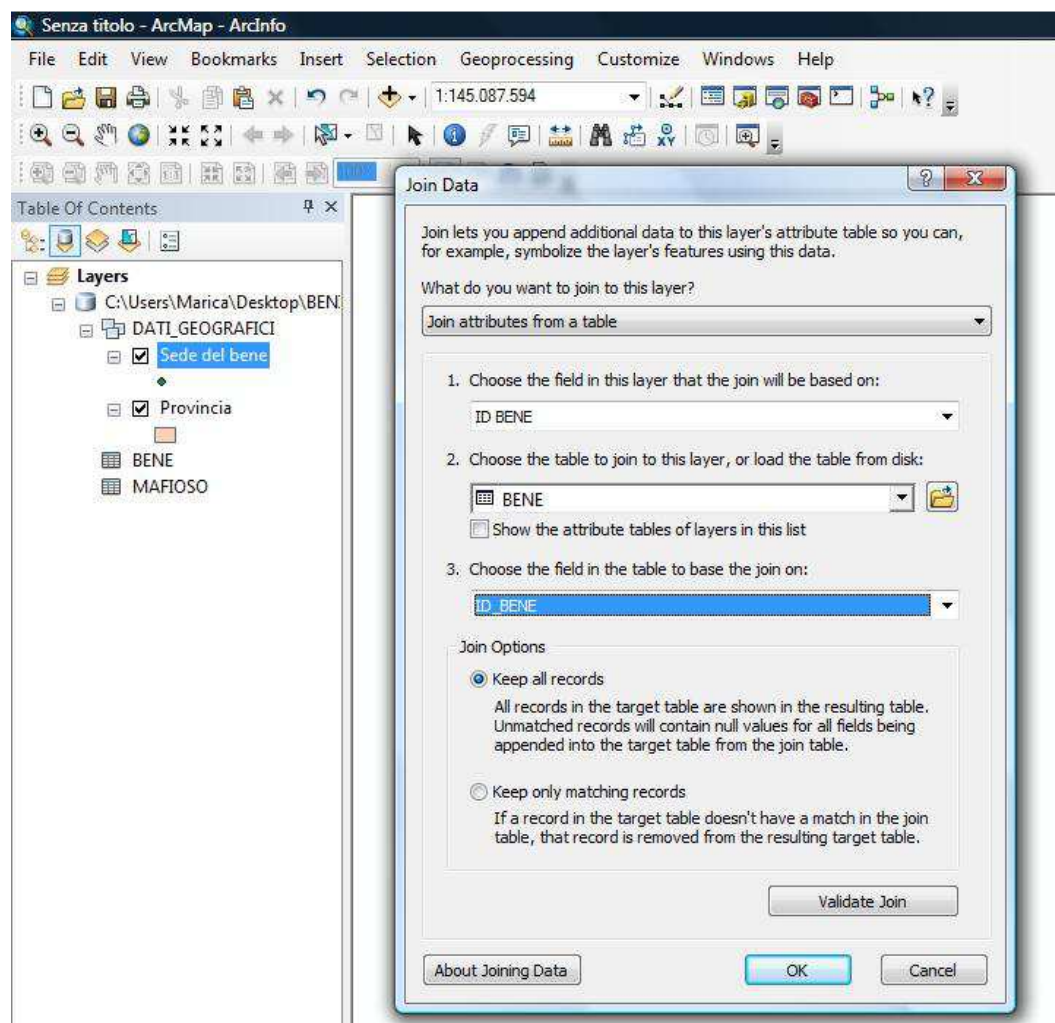


(Figura 3.16. Comando Join and Relate)

³⁶ L'inserimento delle coordinate in un geodatabase spesso avviene importando dati esterni (es. GPS) da memorizzare in tabella. Sono dati che vengono trasformati dal software GIS in *Feature Class* di tipo puntuale: in questo modo, ogni punto verrà posizionato in uno spazio in base alle coordinate che riporta.

³⁷ Come già espresso in precedenza, la funzione *Join* permette di congiungere i record di due tabelle sulla base di un attributo comune.

Si imposti il *Join Data* in modo che gli attributi siano congiunti a partire dalla tabella BENI e che il campo su cui sarà basato il *Join* sia ID_BENE.



(Figura 3.16. Data type e properties della tabella Bene)

Dopo aver cliccato su OK, si noterà che i punti che localizzano i beni forniscono tutte le informazioni presenti nella tabella BENI.

3.8. Interrogazione dei dati geografici

Per spiegare come interrogare i dati presenti nel geodatabase, verrà riportato un esempio di elaborazione statistico-cartografica realizzata a supporto dell'indagine sullo sport e tempo libero, condotta dal Dipartimento Scienze

Giuridiche, della Società e dello Sport, dell'Università degli studi di Palermo. Verrà spiegato come utilizzare il *wizard* di costruzione delle *query* presente nel software open source QGIS.

Ci si è avvalsi di dati tratti da fonti statistiche ufficiali. I dati geografici del sistema della base territoriale regionale sono stati reperiti dal sito dell'Istat; i dati riportanti le percentuali di mutui concessi dall'Istituto per il Credito Sportivo alle regioni italiane, dal 1959 al 2007, sono stati reperiti dal sito dell'ICS³⁸. In particolare, per quanto concerne la base dati cartografica, si è scelto di utilizzare la versione generalizzata dei confini amministrativi, aggiornata al gennaio 2011³⁹. Le unità di riferimento sono le diverse regioni italiane. La variabile riferita a tali unità riguarda l'entità di mutui concessi dall'Istituto per il Credito Sportivo dal 1959 al 2007. Agli aggregati territoriali si è associata una variabile ordinale calcolando la posizione occupata, nella graduatoria dei mutui ottenuti, delle regioni italiane. Il caso riportato è un esempio classico di presentazione dei dati riferiti ad aggregati territoriali, sottoforma di serie territoriale⁴⁰, e descrive come i valori della variabile varino nello spazio.

³⁸ Entrambe le basi dati sono disponibili online e accessibili rispettivamente agli indirizzi <http://www.istat.it/it/strumenti/cartografia> e <http://www.creditosportivo.it/Statistiche.asp?AreaSelezionata=Statistiche&AreaPOS=Link9>

³⁹ Il livello gerarchico regionale in versione vettoriale *shapefile* è ottenuto attraverso una procedura di *dissolve* per aggregazione delle sezioni di censimento a cui viene applicato un algoritmo di semplificazione che consente di ridurre il dettaglio dei poligoni, diminuendone i frastagliamenti. Abbiamo ritenuto di dover usare tale versione poiché più adatta alla creazione di cartografia simbolica.

⁴⁰ Una serie territoriale è una sequenza di valori assunti da una variabile nello stesso momento in diversi aggregati territoriali [Corbetta et al., 69].

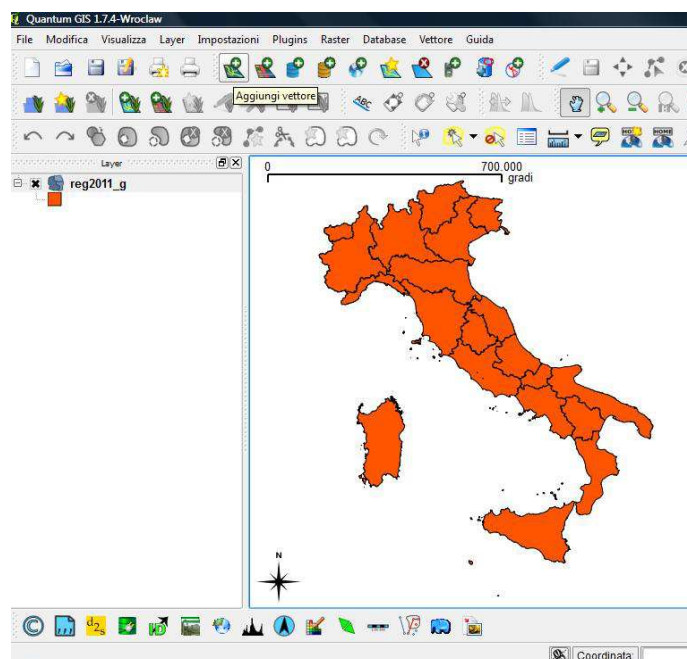
	A	B	C	D
1	REGIONE	N.	IMPORTO	%
2	Valle d'Aosta	36	13.571.320,64	0,24%
3	Molise	227	40.344.685,02	0,70%
4	Basilicata	358	71.846.368,49	1,25%
5	Trentino alto Adige	280	91.374.598,89	1,60%
6	Umbria	632	116.065.073,33	2,03%
7	Sardegna	672	125.287.226,68	2,19%
8	Liguria	604	164.744.879,53	2,88%
9	Abruzzo	643	168.048.677,03	2,94%
10	Marche	1087	190.141.989,25	3,32%
11	Calabria	886	220.830.802,49	3,86%
12	Friuli Venezia Giulia	863	234.900.984,97	4,10%
13	Sicilia	339	264.555.809,23	4,62%
14	Puglia	868	277.487.710,10	4,85%
15	Campania	865	296.041.371,40	5,17%
16	Toscana	1739	408.155.205,50	7,13%
17	Veneto	1824	498.153.458,78	8,70%
18	Piemonte	2139	559.360.901,55	9,77%
19	Emilia Romagna	2107	574.788.627,49	10,04%
20	Lazio	1188	599.844.103,29	10,48%
21	Lombardia	2704	809.958.483,65	14,15%

(Tabella 3.4. Serie territoriale relativa al numero complessivo di mutui concessi dall'ICS, alle diverse regioni italiane, nel periodo che va dal 1959 al 2007)

Verrà costruita una coropleta⁴¹ che raffiguri la distribuzione geografica del fenomeno studiato. Le associazioni e contiguità territoriali, non percepibili attraverso la lettura di una semplice tabella, saranno visibili grazie alla costruzione di una mappa monovariata che faciliti l'analisi comparativa, l'individuazione di omogeneità o di concentrazione di un fenomeno sul territorio [Wright, 1938, 1-18], e la distribuzione di un attributo nello spazio in forma classificata. In tal modo, la misurazione del fenomeno verrà semplificata e la distribuzione territoriale, generalizzata.

Si apra il software QGIS e si carichi lo *shapefile Istat reg2011_g*.

⁴¹ Le carte tematiche di uso più frequente sono le coroplete. Una coropleta rappresenta la distribuzione nello spazio di un attributo di tipo quantitativo e permette di ottenere una descrizione sintetica di come vari.



(Figura 3.18. Caricamento file vettore in QGIS)

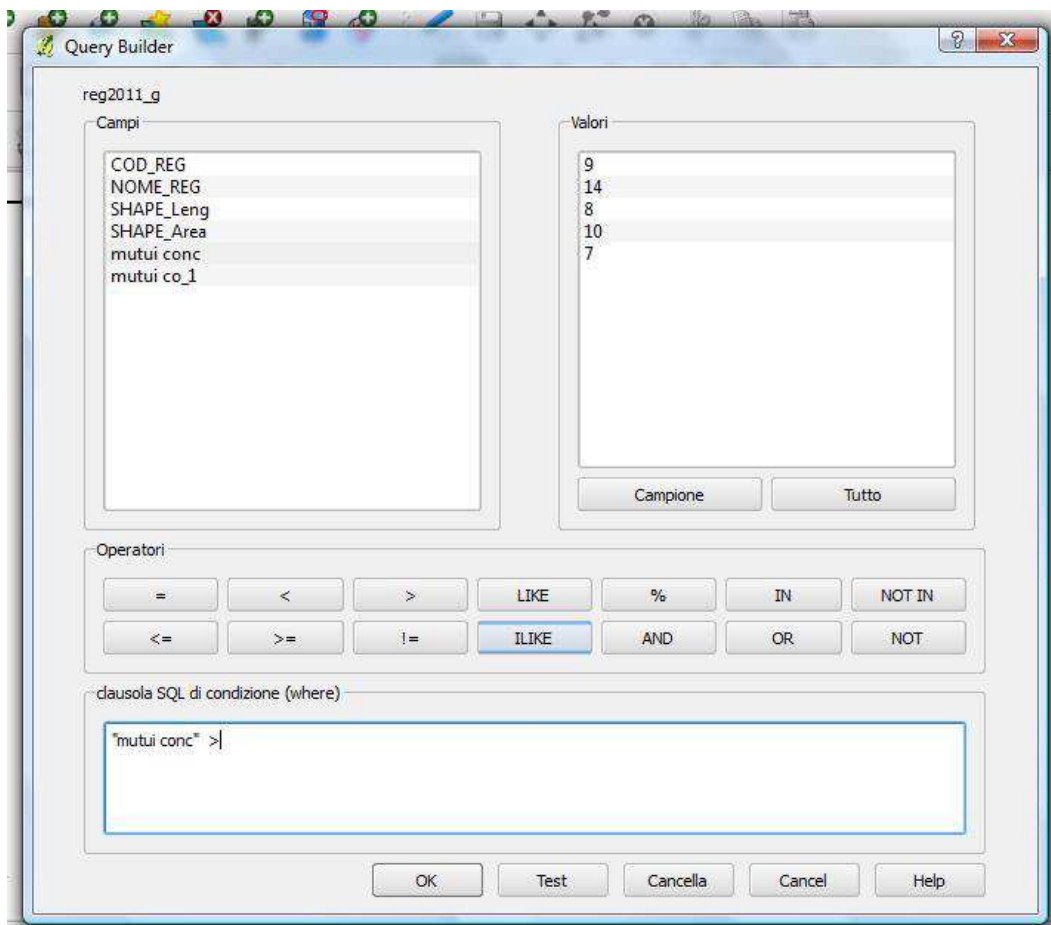
Si clicchi, col tasto destro del mouse, sullo *shapefile* e si scelga il comando *Apri tabella attributi*. Dopo aver cliccato sul comando a forma di penna *Modalità di modifica*, presente nella tabella stessa, si clicchi su *Nuova colonna*. La nuova colonna, da aggiungere alle pre-esistenti e caratterizzanti lo *shapefile*, sarà denominata *Mutui concessi*. Si popoli la colonna coi dati dell'ICS.

	COD_REG	NOME_REG	SHAPE_Leng	SHAPE_Area	mutui concessi
0	1	PIEMONTE	1236869.3688	25394103941.6	559360901.55
1	2	VALLE D'AOSTA...	311165.134877	3259040961.3	13571320.64
2	3	LOMBARDIA	1411265.07152	23862698613.6	809958483.65
3	4	TRENTINO-ALT...	800534.132495	13608018378.5	91374598.89
4	5	VENETO	1057855.9676	18405499479.4	498153458.78
5	6	FRIULI VENEZIA...	667489.737062	7864293935.04	234900984.97
6	7	LIGURIA	834224.468035	5415464956.44	164744879.53
7	8	EMILIA-ROMA...	1164723.05925	22451465521.9	574788627.49
8	9	TOSCANA	1316658.00191	22984426772	408155205.5
9	10	UMBRIA	620315.169929	8464007915.76	116065073.33
10	11	MARCHE	629209.047201	9401177971.95	190141989.25
11	12	LAZIO	1055355.14651	17227617801.3	599844103.29
12	13	ABRUZZO	614513.705789	10829102919.5	168048677.03
13	14	MOLISE	433874.677975	4461149496.36	40344685.02
14	15	CAMPANIA	892379.131913	13663989800.1	296041371.4
15	16	PUGLIA	1176242.09348	19537077223	277487710.1

(Figura 3.19. Tabella attributi del file vettore)

Il passaggio più importante è dato dalla scelta della tecnica più appropriata a raggruppare i dati in classi affinché la coropleta possa essere letta con facilità [Dainelli, 2010]. Per la definizione delle classi dei valori da rappresentare, i GIS offrono funzionalità avanzate. Si potrà decidere di dividere i dati in *classi di pari ampiezza*, in cui la differenza tra il valore massimo e il valore minimo (o campo di variazione) della variabile verrà divisa per il numero degli intervalli scelto. In questo caso, si clicchi, col tasto destro del mouse, sullo *shapefile* e si scelga il comando *Proprietà*. Dalla scheda si clicchi su *Simbolo graduato*, si selezioni la colonna *Mutui concessi*, e si digiti il numero di classi in cui si intende dividere i dati. Oltre che in classi di uguale ampiezza, è possibile dividere anche secondo i metodi *Deviazione standard*, *Raggruppamento naturale* e *Quantili*. Il primo è utile solo se i dati approssimano la distribuzione normale gaussiana e, quindi, solo se questa è caratterizzata da scarsa variabilità e quasi tutti i casi si concentrano attorno alla media. Per applicare il secondo, il software GIS ricerca i punti di rottura o i salti della distribuzione; il risultato è la definizione di classi molto omogenee all'interno, in cui si minimizza la varianza. Il terzo ordina le osservazioni in base alla variabile che si intende rappresentare e le raggruppa in numero uguale di elementi per ogni classe [Boffi, 2004].

È comunque possibile utilizzare altri sistemi statistici e di analisi offerti dai GIS. Cliccando, col tasto destro del mouse, sullo *shapefile* e scegliendo *Query*, sarà possibile invece selezionare i *record* della tabella che rispondono alle *query* e gli elementi geografici legati ai *record*. Si potrà, ad esempio, costruire una *query* che permetta di capire quali siano le regioni in cui si registra una percentuale di concessioni mutui superiore al 10% del totale dei mutui concessi.



(Figura 3.20. Esempio di query)

Come si può notare, nel riquadro *clausola SQL di condizione (where)*, è possibile utilizzare il linguaggio SQL per la costruzione delle *query*. I tasti *Test* e *Cancella* permettono rispettivamente di controllare la sintassi delle espressioni in modo da evitare errori o correggerli, e di ripulire lo spazio di *input*.

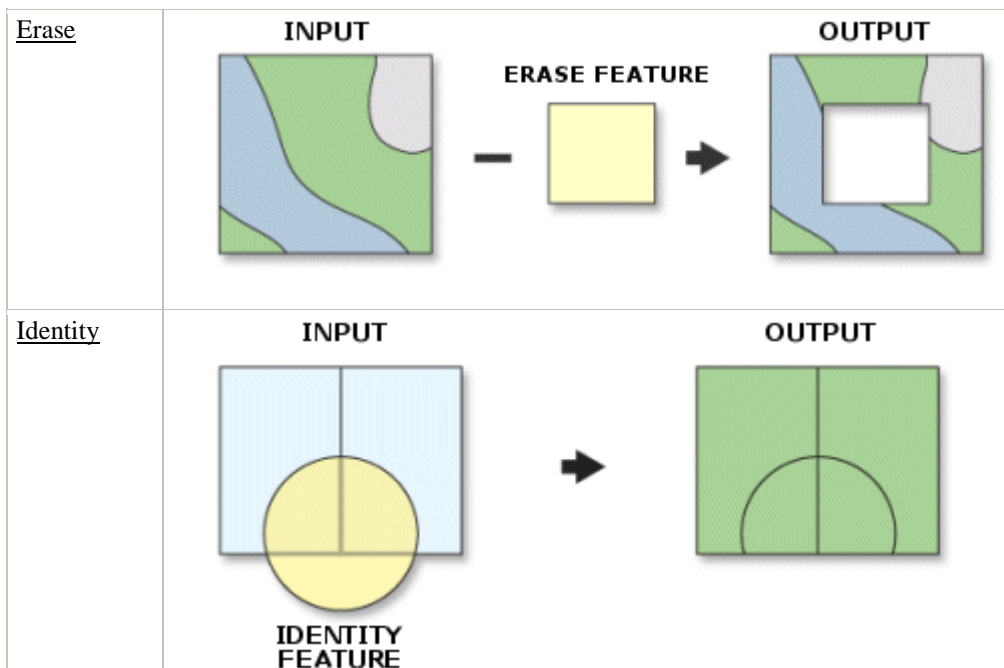
3.9. Overlaying

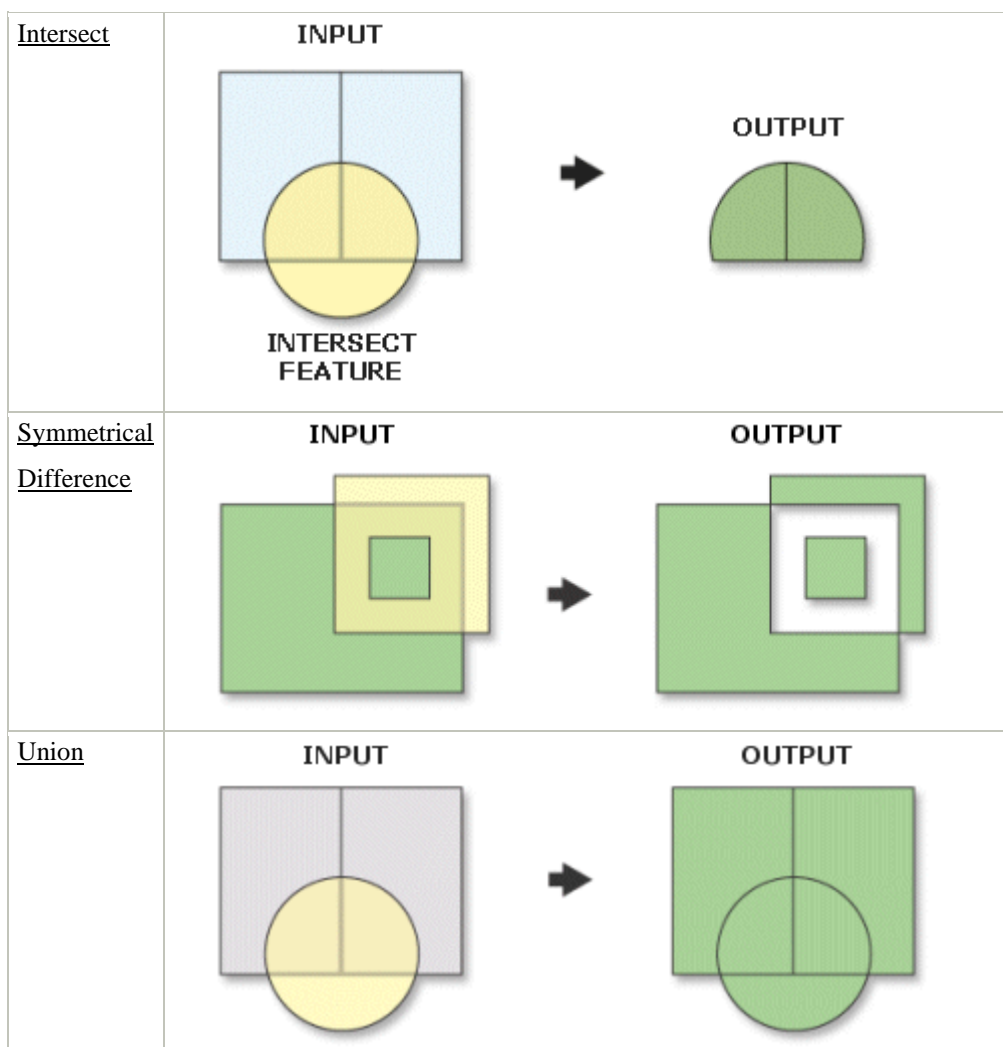
Se, ad un *layer* che riporta la localizzazione degli Starbucks a New York, se ne sovrappone uno che riporta le zone della città in cui la gente parla molto al cellulare, si potrà scoprire che gli oggetti dei due *layer* esprimono una relazione spaziale, affidata alla visualizzazione e all'elaborazione intuitiva [Shi *et al.*, 2007]. Si tratta della tecnica multistrato, o *overlay*, che permette il confronto tra strati informativi spaziali diversi mettendo in relazione oggetti dello stesso tipo o

di tipo diverso. Permette di studiare analiticamente le relazioni nello spazio in base alle proprietà geometriche degli oggetti.

Oltre all'operazione di *join*, di cui abbiamo già discusso, quelle più importanti di *overlay* sono [ESRI, 2013]:

- *Erase*, che genera una nuova *Feature Class* sovrapponendo la *Feature di input* con il poligono della *Erase Feature*, in modo che solo le porzioni di spazio appartenenti alla *Feature di input*, e che rimangono al di fuori dai confini della *Erase*, vengano copiati nella nuova *Feature Class*;
- *Identity*, che genera una *Feature Class* includendo nella *Feature di input* le porzioni comuni con l'*Identity Feature*;
- *Intersect*, che genera, come *output*, una *Feature* che comprende solo l'area d'intersezione tra la *Feature di input* e quella d'intersezione;
- *Symmetrical Difference*, che esclude dalla *Feature di output* la differenza tra la *Feature di input* e la *Feature di intersezione*;
- *Union*, che unisce due *Feature Class*;





(Figura 3.21. overlaying. Fonte: ESRI Italia)

A parte l'operazione di *Join*, che lascia invariati gli attributi della *Feature di input* [Parmenter, 2007], lo scopo dell'*overlaying* degli oggetti geometrici contenuti in mappe vettoriali, genera nuovi attributi che caratterizzeranno il nuovo strato. Le operazioni di *overlaying* effettuate sui *layer* Beni di Sicilia sono tra le più comuni. Ad esempio, la sovrapposizione della distribuzione dei punti (i beni localizzati in base alle coordinate geografiche) sui poligoni rappresentanti aree amministrative (le Province) o il conteggio del numero di beni presenti nell'area. Sono operazioni che è possibile effettuare costruendo le *query* e attraverso gli operatori logici spaziali in grado di calcolare le relazioni tra gli oggetti [Perencsik *et al.*,

2004]. Sono strumenti che studiano le relazioni spaziali, enumerandole secondo diversi criteri topologici e che individuano aree caratterizzate dalla compresenza di più fenomeni.

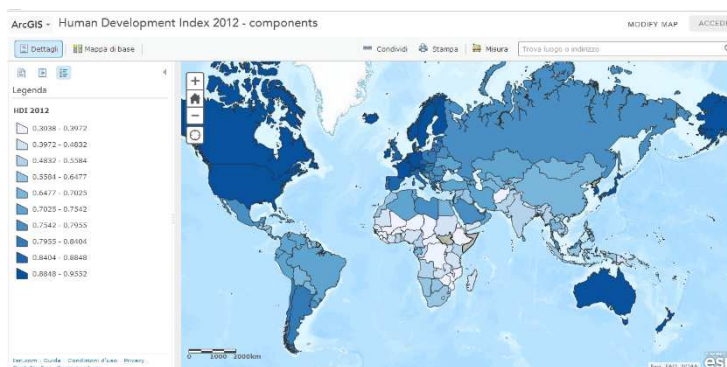
Come già spiegato nel corso del Capitolo Secondo, l'*overlaying* tra mappe *raster* prevede che le celle, di cui si compongono le griglie appartenenti a uno o più *layer*, possano essere sovrapposte tra loro. Le operazioni che è possibile effettuare tra i due strati continui sono le aritmetiche e le logiche.

Se la sovrapposizione viene effettuata tra strati che rappresentano ognuno un attributo diverso, il *layer* di *output* rappresenterà l'indicatore sintetico nello spazio [Travaglini, 2004].

3.10. Overlaying con ArcGIS online

Con la diffusione di dataset di diverso tipo, anche i geodata stanno cominciando a diffondersi. Molti di questi sono resi disponibili attraverso le versioni online dei software GIS. Uno di questi è ArcGIS online, accessibile attivando un account gratuito⁴². Si riportano alcuni esempi di utilizzo.

Si clicchi nel box di ricerca in alto a destra e si digiti *Human Development Index 2012 – components*. Tra i diversi risultati che appariranno, si clicchi sulla mappa edita da *edugiscoe*.



(Figura 3.22. Human development Index in ArcGIS online)

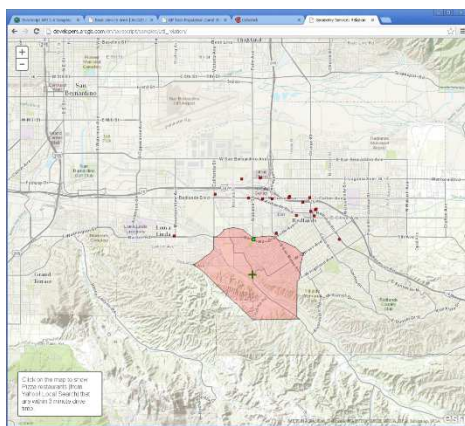
Si clicchi sul pulsante *Mostra il contenuto della mappa*, presente in alto a sinistra, nella tabella dei contenuti. Si punti il *box* accanto a HDI 2012 e, tra quelle offerte

⁴² <http://www.arcgis.com>

dal menu a tendina, si scelga l'opzione *Cambia simboli*. Sarà possibile scegliere il numero delle classi e le modalità attraverso le quali ripartirle. Cliccando ancora sulla freccia accanto a HDI 2012 e selezionando *Mostra tabella*, si potrà avere accesso alla tabella degli attributi. Per interrogarla, è possibile utilizzare la sezione *Opzioni tabella* e scoprire, attraverso il filtro, quale paese sia caratterizzato, ad esempio, da un HDI più basso di un certo valore scelto. Dalla finestra di dialogo alla quale è possibile accedere attraverso la sezione *Filtro*, si selezioni *Human Development Index (HDI) 2012*. Dal box immediatamente a destra, si selezioni l'opzione *è minore di e*, nel campo *valore*, si digiti ,5. Infine si clicchi su *Applica filtro*.

In questo modo, sarà possibile visualizzare, sulla nostra mappa, solo quei paesi che hanno un HDI minore di 0,5. Per interrogare la mappa e impostare altri filtri, basta rimuovere il filtro applicato.

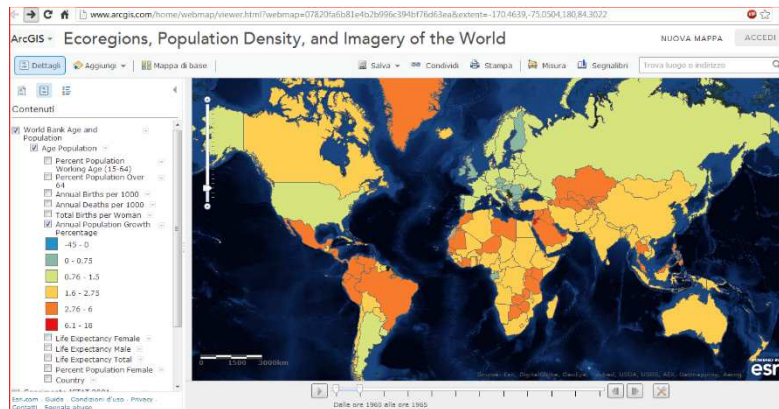
Con ArcGIS *online*, è inoltre possibile computare i dati registrati nelle griglie *raster*. Si digiti, ad esempio, l'indirizzo: http://developers.arcgis.com/en/javascript/samples/gp_zonalstats/.



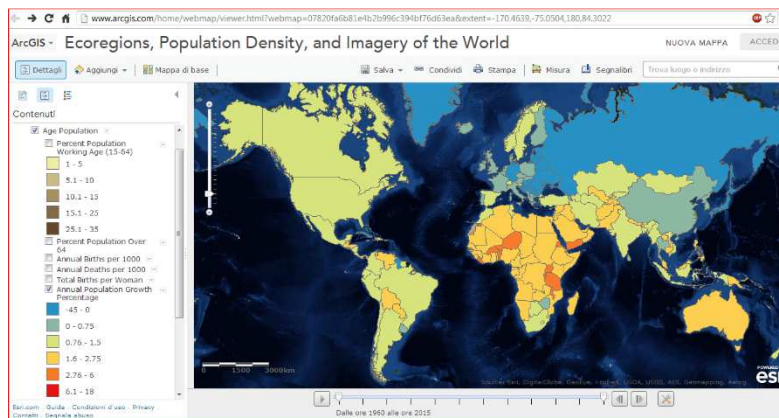
(Figura 3.23. Il poligono disegnato comprende il territorio delle città di Atlanta e Georgia)

Cliccando su *Summarize Population*, si disegni un poligono. Il *tool* restituirà un risultato che riporta il numero di abitanti residenti nella zona delimitata dai lati del poligono.

È possibile agire, inoltre, sulla variabile tempo. Si digiti l'indirizzo: <http://www.arcgis.com/home/webmap/viewer.html?webmap=07820fa6b81e4b2b996c394bf76d63ea&extent=-170.4639,-75.0504,180,84.3022>. Si avrà accesso alla mappa delle eco-regioni i cui *layer* sono stati prodotti da *Intl_User_Community*.

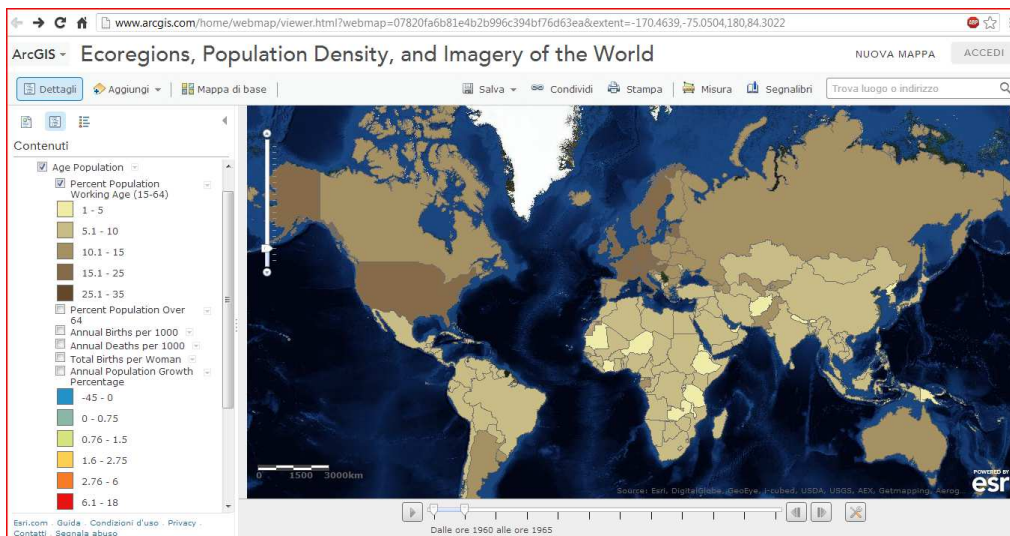


(Figura 3.26. Ecoregioni, densità della popolazione considerato l'arco temporale che va dal 1960 al 1965)



(Figura 3.27. Ecoregioni, densità della popolazione considerato l'arco temporale che va dal 1960 al 2015)

Si accenda il *layer Percentuale popolazione in età lavorativa (15-64)* e si scelga l'arco temporale entro il quale si è interessati a verificare come si evolve tale distribuzione.



(Figura 3.28. Ecoregioni, densità della popolazione. Layer riportante dati sulla percentuale di popolazione in età lavorativa dal 1960 al 1965)

Si potrà notare, ad esempio, che dal 1960 al 1965, gli Usa e i paesi europei sono tra quelli con una percentuale più alta di popolazione in età lavorativa (dal 15 al 25%), mentre alcuni dei paesi africani come la Nigeria o la Mauritania hanno la percentuale più bassa (dall'1 al 5%).

Ora, è chiaro come il GIS online semplifichi la consultazione degli archivi di dati e che avere accesso ai dati messi a disposizione dalle grandi case proprietarie di software, spesso, rappresenta l'unico modo per poterli consultare per le proprie analisi. Bisogna, però, sottolineare che la differenza tra la costruzione *ex novo* del *layer* e l'utilizzo del *layer* preconfezionato sta nel limite generato dall'utilizzo del secondo e dalla conseguente scarsa interrogabilità del geodatabase. Quando l'analista è in grado di costruire un geodatabase da sé, riceverà le risposte più appropriate alle *query* poste [Steinberg e Steinberg, 2006]. Esse saranno frutto del suo stesso disegno progettuale.

CAPITOLO QUARTO GEODATA

Definito il disegno della ricerca, l'analista GIS dovrà affrontare la delicata fase che riguarda la selezione del dataset appropriato all'analisi. Non si ribadisce mai a sufficienza che le tecniche di ricerca sociale non migliorano la qualità dei dati e che questa pone precisi vincoli alla bontà dei risultati ottenuti. Perciò, oltre a quelle già elencate nel corso della presente trattazione, le prime domande che l'analista dovrà porsi, nel momento in cui si apprestia ad applicare le tecniche GIS, sono le seguenti: «Esistono dati utili alla mia ricerca, in formato compatibile col software a mia disposizione?»; «Se sì, ho accesso a questi dati?»; «A quali condizioni?»; «Chi li ha raccolti originariamente?»; «A che scopo?»; «Quali attributi sono inclusi nel dataset?»; «Sono dati recenti?».

Si è già dimostrato che costruire un geodatabase *ex novo* è un'operazione complessa. Considerato il tempo che occorre per produrre autonomamente dati utilizzabili in un GIS, lo sforzo di cercare dati esistenti e impiegare le competenze e abilità tecniche possedute per adattarli all'indagine può essere affrontato senza timore dal ricercatore. Nell'utilizzo dei GIS, infatti, è racchiuso il nucleo di una prospettiva la cui rivalutazione sta già segnando un passo nuovo nell'ambito della metodologia della ricerca sociale. Si tratta dell'approccio basato sull'analisi di dati esistenti, un vero e proprio modo di fare ricerca, meno oneroso e altrettanto valido scientificamente [Biolcati-Rinaldi e Vezzoni, 2013].

Qui, non ci si rivolge al ricercatore sociale come se questi si trovasse nel bel mezzo del deserto conoscitivo, costretto a dover affrontare questioni a cui nessun altro ha mai pensato, a rilevare dati mai disponibili e che nessun altro ha mai prodotto o rilevato. Piuttosto, coerentemente con quanto avviene nella realtà, riteniamo che il percorso di ricerca GIS non debba essere ingessato in un modello tipico e che la cumulatività delle informazioni disponibili favorisca la loro fruizione creativa.

La facilità con cui i geodata vengono prodotti e accumulati con elevata frequenza, li connota come un patrimonio informativo insostituibile per lo studio delle stesse comunità che li producono. La loro analisi costituisce un'opportunità sia per la verifica e la revisione di asserti prodotti da altri osservatori GIS, che per la risposta a domande non individuate e/o a questioni non ancora affrontate [Macintyre *et al.*, 2002; Openshaw, 1984]. L'analisi dei geodata secondari ha il potenziale di rendere fallibile la GIS *analysis* primaria altrui o, meglio ancora, la propria. È altamente desiderabile, infatti, per il bene della scienza sociale, che l'onesto ricercatore metta in discussione i suoi stessi tentativi congetturali e le sue falsificazioni, diffondendo i geodata originali, indicando ad altri quali sono gli strumenti per aggregarli o modificarli.

Mettere insieme dati provenienti da diverse fonti può essere complicato. Ma, il non lasciarsi scoraggiare, ripagherà certamente l'analista permettendogli di incorporare informazioni eterogenee, normalmente difficili da integrare tra loro, in un ambiente di analisi unico.

A tal proposito, in questo paragrafo, si farà riferimento al modo in cui egli potrà trovare geodata già prodotti e raccolti, estrarli dalle loro miniere inesauribili adattandoli alle proprie esigenze conoscitive [Steward, 1984]. Non è certo che sceglierli porti necessariamente alla verità; probabilmente, nemmeno alla certezza di aver raggiunto la vetta desiderata ma può rappresentare un ingente risparmio e, spesso, l'unica possibilità di studiare il problema in maniera comparata migliorando l'attendibilità dei geodata stessi.

Il non imbrigliarsi nel sistema dei dati, ma percorrerlo senza lasciarsi immobilizzare o dominare, è questione di metodo e di esercizio pratico. Occorre scegliere i geodata in maniera pertinente mantenendo saldo il principio dell'indissolubilità tra teoria e ricerca, tra metodo e tecniche.

Stabilire se i geodata secondari siano appropriati o meno alla nostra ricerca è importante tanto quanto preoccuparsi che siano di qualità. I dati empirici da analizzare vanno sempre adeguati all'obiettivo e al tipo di analisi da compiere [Zajczyk, 1996] ma occorre, prima di tutto, che siano costituiti da due componenti essenziali:

- Il corredo geografico, attraverso il quale geolocalizzare gli oggetti nello spazio considerato;
- L'insieme di attributi, utile a descrivere le proprietà delle entità o oggetti spaziali.

È, inoltre, auspicabile che i dataset contengano le informazioni accessorie che individuano le relazioni topologiche tra le entità.

Un buon analista GIS è in grado di capire come i geodata siano stati generati, chi li ha prodotti e quali siano i limiti posti dalla loro natura. Sa sceglierli in base ai suoi obiettivi, alla sua esperienza, alle risorse che ha a disposizione e al livello di diffusione dei dati stessi.

Il problema della scarsa diffusione pubblica dei geodata non è irrilevante poiché potrebbe influenzare l'intero disegno o alterare la domanda di ricerca. Perciò, occorre valutare con attenzione se il dataset disponibile sia appropriato o meno a rappresentare le relazioni spaziali, in base al livello di dettaglio scelto. Ad esempio, se l'unità di analisi è il quartiere e i dati a disposizione riportano informazioni su un'unità più ampia, se il dataset non riporta informazioni aggiuntive che permettono di capire come e quando i dati siano stati raccolti e da chi, allora sarà conveniente cercare un nuovo set assicurandosi che il livello di aggregazione sia adeguato agli obiettivi posti dalla ricerca.

4.1. Trovare i geodata

Quando si compie un'analisi secondaria, in quanto applicata a informazioni ottenute in precedenza da altri, occorre chiedersi quali istituzioni li renda disponibili e in base a quali fini. Oggi, la fruibilità dei geodata è agevolata dall'implementazione dei cosiddetti *Clearing-house* o archivi di dati online [Biolcati-Rinaldi e Vezzoni, 2013]. L'idea su cui si fondano è quella secondo la quale non esiste un unico modo ottimale di accedere ai geodata; piuttosto, esistono tanti modi differenti adatti alle esigenze dell'utente. Egli può navigare tra i dati scegliendo, tra i percorsi di consultazione, quelli che ritiene più pertinenti ai suoi interessi, selezionando anche il livello di dettaglio più appropriato

all'indagine. Un esempio è il *DaWinci*⁴³ predisposto dall'Istat sui censimenti del 2001. L'analista può selezionare i dati in base a tre parametri: l'oggetto di interesse (ad esempio, i praticanti sport nel 1985), le classificazioni secondo cui si vogliono estrarre i dati relativi all'oggetto prescelto (ad esempio, genere, età, ecc.) e il territorio di riferimento (ad esempio, la regione Sicilia). Il sistema di cartografia interattiva, di cui è dotato tale *clearing-house*, permette inoltre di visualizzare cartogrammi tematici.

I *clearing-house* possono essere costituiti da archivi informatici, decentrati in varie parti del mondo e gestiti da enti diversi. Le interrogazioni avvengono accedendo al sistema della rete informativa geografica, dotato di una guida utile alla definizione dei criteri di ricerca.

Esistono milioni di *provider* di dati online che si dividono in quattro gruppi: università, enti di ricerca pubblici, organizzazioni non profit, istituti di ricerca privati. Essi mettono a disposizione dell'analista un immenso patrimonio informativo a cui attingere per diversificate indagini.

Tra gli archivi più conosciuti:

- Il *Census Bureau*⁴⁴ che, negli Stati Uniti, offre l'accesso a dati nazionali e internazionali nell'area socio-demografica;
- Il CIESIN⁴⁵ (*Center for International Earth Science Information Network*), della Columbia University, che archivia dati riferiti a diverse aree del mondo, con particolare attenzione ai temi delle scienze sociali;
- AI-GEOSTATS⁴⁶, curato dal *Joint Research Center* della Commissione Europea;
- GloVis (*Global Visualization Viewer*)⁴⁷ della USGS (*United States Geological Survey*);
- GNIS (*Geographic Names Information Systems*)⁴⁸ che riporta i dati relative ai luoghi degli stati confederati dei territori USA;

⁴³ <http://dawinci.istat.it>

⁴⁴ <http://www.census.gov/>

⁴⁵ <http://www.ciesin.org/>

⁴⁶ <http://www.ai-geostats.org>

⁴⁷ <http://glovis.usgs.gov/distribution>

⁴⁸ <http://www.usgs.gov>

- GNS (*GEOnet Names Server*)⁴⁹ della NIMA (*National Imagery and Mapping Agency*) che riporta i dati geografici relativi a tutti gli stati del mondo.

Le fonti di geodata più attendibili in Italia sono principalmente quelle dirette a livello ministeriale. Gran parte del lavoro di produzione veniva affidato, sino a qualche anno fa, quasi esclusivamente all'IGM (Istituto Geografico Militare) poiché le mappature erano prodotte principalmente per scopi di difesa, pianificazione e controllo del territorio. Nonostante il lavoro dell'IGM abbia rappresentato, per lungo tempo, un punto di riferimento per cartografi e analisti, i *topic* di analisi oggi sono assai diversi e possono essere difficilmente affrontati esclusivamente attraverso dati che comprendono poche tipologie, quali mappe catastali su grandi scale, mappe topografiche su media scala e nazionali su piccola scala.

Oggi, l'Istat produce geodata riferiti ai confini amministrativi (Regioni, Province, Comuni) e basi territoriali in formato *shapefile*. Ha previsto, tra i suoi progetti, anche l'integrazione progressiva degli archivi e dei sistemi informativi, utilizzando una base geografica comune, allo scopo di sviluppare un nuovo sistema cartografico di riferimento [Istat.it].

Oltre a Istat e IGM, esistono numerosi altri enti istituzionali che raccolgono, elaborano, gestiscono informazioni statistiche cartografiche [Boffi, 2004]: il SISTAN (Sistema Statistico Nazionale), le Amministrazioni dello Stato (tra i Ministeri, i Dipartimenti e le Aziende autonome, da citare l'Istituto Superiore per la Protezione e la Ricerca Ambientale e il Ministero dell'Ambiente, le cui banche dati sono particolarmente ricche e aggiornate), gli enti territoriali (Regioni, Province Autonome).

Tra gli enti pubblici, ACI (Automobile Club d'Italia), CISPEL (Confederazione Italiana Servizi Pubblici Enti Locali), CNEL (Consiglio Nazionale dell'Economia e del Lavoro), ENEL (Ente Nazionale Energie Elettrica), ENIT (Ente Nazionale Italiano per il Turismo), INAIL (Istituto Nazionale per l'Assicurazione per gli

⁴⁹ <http://earth-info.nima.mil/gns/html/index.html>

Infortuni sul Lavoro), INPS (Istituto Nazionale della Previdenza Sociale), SIAE (Società Italiana degli Autori e degli Editori), RAI (Radio Televisione Italiana), Tribunali, Distretti Scolastici, Distretti Militari, ecc.

Tra gli enti scientifici, di particolare rilevanza il ruolo svolto da CNR (Consiglio Nazionale delle Ricerche) e da tutte le Università e le Fondazioni che elaborano autonomamente i dati mettendoli a disposizione della comunità scientifica.

Tuttavia, quando si ha intenzione di condurre un'analisi su grande scala e di restringere, quindi, l'area di interesse a quella locale, le risorse gratuite online tendono a soddisfare poco le esigenze del ricercatore. In questo caso, potrebbe essere utile rivolgersi ai *provider* di dati online accessibili a pagamento o alle fonti di dati offline [Steinberg e Stienberg, 2006].

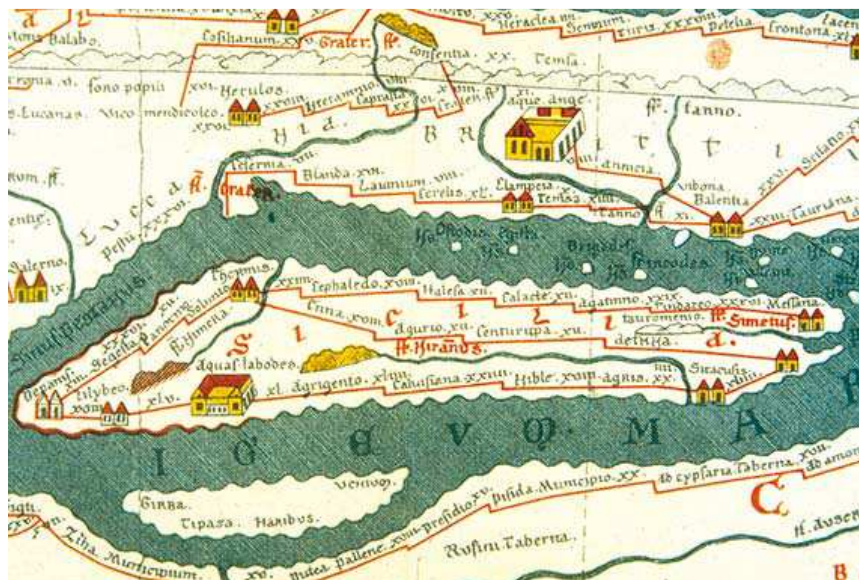
Buona parte dei GIS data riferiti al territorio italiano, e venduti dagli istituti di ricerca privati, non sono altro che elaborazioni alternative di dati messi a disposizione dagli enti di ricerca pubblici. Sono dati, spesso, più semplici da utilizzare e compatibili con i software GIS più diffusi [Atzeni *et al.*, 2003]. Tuttavia, il ricercatore dovrà valutare se il costo di accesso potrà essere ammortizzato dal beneficio dato dall'utilizzo di tali dati preconfezionati o se, piuttosto, avendo a disposizione abbastanza tempo, sia preferibile per lui scegliere il dataset originale nel quale integrare autonomamente le informazioni desiderate.

Le fonti offline possono offrire delle soluzioni alternative [Biolcati-Rinaldi e Vezzoni, 2013]. Esistono numerosi archivi storici ai quali è possibile accedere. Anche in questo caso, occorre verificare l'attendibilità e la qualità del dato e tenere in considerazione che, nel caso di dati provenienti da archivi storici, è probabile che esistano solo in formato cartaceo e che non siano stati digitalizzati. Ogni dato esistente in formato fisico richiede la conversione in formato digitalizzato per essere sfruttato in un'analisi GIS. Il primo passo da compiere nel processo di digitalizzazione è l'inserimento manuale dei dati nel computer.

Di recente, data l'ingente quantità di risorse necessarie a digitalizzare dati in formato chiuso e i benefici prodotti dal processo di conversione, associazioni

come Spaghetti Open Data⁵⁰ hanno deciso di svolgere volontariamente questo compito utile alla collettività e alla ricerca. I vantaggi derivanti dall'utilizzo di dati storici è evidente: se comparabili, potrebbero essere utilizzati come informazione utili a monitorare *trend* e cambiamenti nel tempo. La difficoltà sta proprio nella digitalizzazione di dati comparabili ad altri esistenti, prodotti in diversi momenti storici e da diversi soggetti.

Si immagini di dover comparare la rappresentazione cartografica della Sicilia riportata nella *Tabula Peutingeriana*⁵¹, eseguita alla fine del IV secolo, con un'immagine satellitare [Allen, 1993].



(Figura 4.1. *Tabula Peutingeriana*)

Riscontreremmo differenze sia per quanto riguarda la qualità dei dati che per la forma e le dimensioni degli oggetti geografici riportati [Prontera, 2003]. Le

⁵⁰ <http://www.spaghetiopendata.org/>

⁵¹ La crescita della dominazione romana costrinse i geografi ad estendere le rilevazioni a territori sempre più lontani e a stabilire i cosiddetti itinerari che avevano finalità esclusivamente di carattere militare e che possono oggi essere definiti delle carte stradali ante-litteram. La rappresentazione in questione, eseguita ai tempi di Teodosio Magno alla fine del IV secolo, contiene i lineamenti delle terre e della rete stradale dell'Impero Romano, dalla Penisola Iberica al Mar Caspio. In tale proiezione grafica la rappresentazione della Sicilia è deformata, compressa, consentendo la consultazione degli itinerari in formato portatile.

tecniche di rilevazione dei dati sono estremamente diverse, così come le finalità che hanno condotto alla produzione di entrambe le mappe e dei dati ad esse correlate.

In genere, i *layer* cartografici offerti dai principali *provider*, istituzionali e non, riportano i tipi di *feature* più comuni come i confini politici o le principali caratteristiche relative al paesaggio. In base al livello di dettaglio richiesto dallo studio, potrebbe essere necessario ricorrere a mappe a grande scala che contengano anche informazioni riportanti gli indirizzi e i numeri civici, i confini dei quartieri. Una mappa così dettagliata semplifica il processo di compilazione del database con informazioni aggiuntive [*Ibidem*]. Ad esempio, ad una mappa di una certa area urbana, in cui sono riportati i dati riferiti a tutte le abitazioni dotate di un telefono fisso, si possono collegare dati riguardanti i rispondenti a un questionario. Per collegare i dati raccolti a una mappa esistente occorre però interpretare il modo in cui la mappa è stata codificata. Conoscere i codici degli oggetti riportati facilita il processo di codifica dei dati raccolti *ex novo* e che dovranno essere aggiunti al geodatabase [National Research Council, 2007].

Insomma, potrebbero volerci dei mesi per raccogliere i dati che servono, specie se si tratta di dati non digitalizzati esistenti solo in formato cartaceo. E anche quando i dati digitalizzati sono disponibili, è possibile che occorra manipolarli perché possano essere integrati nel GIS che vorremmo costruire [Prontera, 2003]. Certo è che, nonostante le evidenti difficoltà nel loro reperimento, occorre ricordare che trovare i dati da usare in un'analisi GIS non è molto diverso dal trovare quelli necessari ad una qualsiasi altra analisi di ricerca sociale. I dati non sono disponibili immediatamente quando il ricercatore ne ha bisogno. Per questo, occorre che questi investa buona parte del suo tempo e delle sue risorse nella fase dedicata alla loro raccolta.

È, inoltre, necessario che anche i ricercatori, oltre che i professionisti e gli sviluppatori, diano il loro contributo mettendo a disposizione di altri i dati da loro prodotti. A tal proposito, la produzione e diffusione dei dati GIS da parte dei ricercatori sociali genera un circolo virtuoso che favorisce il loro utilizzo da parte di chi intende corroborare i risultati ottenuti da altri o metterli in discussione. Del

resto, una ricerca sociale sempre più *cross-national*, e che manifesta chiare difficoltà legate al reperimento dei finanziamenti, necessita d'investire su cambiamenti sostanziali a livello organizzativo e di formare impresa collettiva interessata ai dati come bene condiviso [Biolcati-Rinaldi e Vezzoni, 2013].

4.2. Metadata

Inserire i *metadata* nel dataset significa attribuire un senso alle quantità desunte dalle fonti statistiche e utilizzate nell'ambito di ulteriori indagini [Mingo, 2007]. I *metadata* sono informazioni sulle informazioni che indicano come i dati sono stati prodotti, per quale scopo, quando e da chi. Perciò, quando i dati sono corredati dai *metadata*, determinare la loro adattabilità allo studio in corso è certamente più semplice. Tra i *metadata*, è possibile ritrovare anche informazioni accessorie riguardanti il formato, il sistema di proiezione usato, il sistema di coordinate, il *Datum* e il software con il quale è possibile modificare i dati. Complessivamente, esistono oltre 300 tipi di *metadata* che possono essere associati al dataset nella fase di costruzione del geodatabase⁵². È indispensabile usare quelli che veicolano informazioni esaustive sui dati, in modo che l'utente non sia costretto ad effettuare il download e ad aprire il dataset su un software desktop GIS per verificarne il contenuto [Berners-Lee *et al.*, 2006]. I *metadata* permettono di individuare il campo di indagine, il contesto spazio-temporale in cui i dati sono stati rilevati e che riguarda sia la popolazione di riferimento che il periodo di rilevazione. Proprio l'informazione temporale è una tra le più importanti da inserire nei *metadata*; in particolare, quella che riguarda l'aggiornamento dei dataset. Per quelli che sono composti da dati riferiti a fenomeni che registrano cambiamenti repentini e che, quindi, sono caratterizzati

⁵² L'utilizzo dei metadati non è molto diffuso tra i GIS esperti operanti nella Pubblica Amministrazione. Si riscontrano, con una certa facilità, casi in cui i dataset resi disponibili agli utenti sono corredati da informazioni insufficienti a capire quali scopi abbiano animato la produzione del dato. Ciò li rende, ovviamente, inutilizzabili e non modificabili. Riguardo il dataset delle aree verdi del Comune di Bari (<http://opendata.comune.bari.it/>), non si è in grado di conoscere a priori il tipo di primitiva geometrica contenuta, gli attributi associati agli elementi geografici ed il loro significato. Solo aprendo i file contenuti nel .zip con un desktop GIS, si scopre che i dati geografici sono composti da poligoni. Se poi si interroga un poligono compare una lista di 34 attributi che, per la maggior parte, riportano nomi incomprensibili.

da una forte variabilità nel tempo (ad esempio, i flussi migratori), è importante specificare anche la frequenza di aggiornamento. Infatti, il ritardo con cui i dati vengono resi disponibili, rispetto al momento della rilevazione, può renderli non attuali rispetto agli interessi conoscitivi degli utilizzatori [Belussi, 2006].

Le informazioni che consentono di identificare la popolazione di riferimento consentono, invece, di capire se l'indagine si riferisca a grandi o piccole quantità di unità, se è su grande o piccola scala, ecc.

Un altro tipo di *matadata* è quello che riferisce sul tipo di indagine, la sua estensione e le modalità di rilevazione. È utile informarsi sul tipo di campione usato e sulla sua numerosità e sulle modalità di rilevazione. Al fine, poi, di codificare adeguatamente i propri dati in modo che siano compatibili con quelli secondari, occorrono informazioni riguardo le definizioni esplicite o operative di ogni variabile. Infatti, il sistema di attribuzione del codice non univoco può generare incomparabilità tra i dati secondari e i propri da incorporare nello stesso geodatabase [Berners-Lee *et al.*, 2006].

Insomma, i *metadata* consentono di delineare il significato dei dati, di stabilire se quelli ai quali si attinge sono adeguati o meno, congruenti o meno con le specifiche finalità dell'indagine nell'ambito della quale l'analista GIS vuole utilizzarli. Perché rispondano allo scopo di una maggiore usabilità dei dati stessi, alcuni organismi nazionali e internazionali tentano di codificare uno schema concettuale universale dei dati e *metadata*. Un comune linguaggio di descrizione venne definito per la prima volta negli Stati Uniti, nel 1994, a cura del FGDC (Federal Geographic Data Committee): il *Content Standard for Digital Geospatial Metadata*, adottato oggi a livello internazionale.

4.3. Accuratezza e precisione dei geodata

I termini accuratezza e precisione sono spesso usati per distinguere fra errori sistematici e casuali: si dice che una misura è accurata quando contiene piccoli errori sistematici; è precisa quando contiene pochi errori casuali [Corbetta, 2001]. Un conteggio di poche unità può fornire una misura precisa e, con grande probabilità, la sua ripetizione determina lo stesso valore. Un conteggio ripetuto di

un campione numeroso difficilmente conduce allo stesso risultato per la frequenza con la quale si possono commettere errori [Castino e Roletto, 1991]. Quando si dispone di misure ripetute, la distribuzione dei valori può essere rappresentata e quantificata mediante gli indici della statistica descrittiva. Essi servono per rispondere a due domande: «Quale è il valore reale del fenomeno?»; «Come descrivere la variabilità del fenomeno o l'errore commesso nella sua misura?». Al momento della raccolta dei dati, occorre quindi tenere presente che i valori devono essere misurati con la precisione utile a fornire una risposta accurata alle due domande precedenti.

Tuttavia, a causa della grande variabilità dei dati ambientali e territoriali, queste misure non conducono mai a risultati identici. Il divario che separa inevitabilmente i geodata che vengono registrati nel geodatabase dalla realtà che si intende studiare determina l'inaccuratezza del geodato stesso. È una questione strettamente connessa a quella della qualità del dato [Corbetta *et al.*, 2001]. Nel caso dei GIS, tale questione va di pari passo con la loro capacità di integrare nel geodatabase tipi diversi di dati. Ogni *layer* informativo porta con sé errori inevitabili che si combinano con gli errori presenti nei dati originari. L'accuratezza è certamente direttamente proporzionale alla grandezza di scala di una mappa: è inevitabile che, per la costruzione di una grande scala, venga richiesta maggiore accuratezza e precisione. Tuttavia, occorre considerare anche le diverse fonti di errori che influenzano tale precisione e accuratezza [Boffi, 2004, 121-122]:

- Mancata sincronizzazione temporale tra i geodata e gli attributi;
- Utilizzo di geodata a una scala maggiore di quella consentita dalla precisione di rilevazione;
- Copertura areale disomogenea nelle diverse zone e metodologie diverse utilizzate per la rilevazione del dato;
- Trasformazioni sul dato geografico (ad esempio, il cambiamento della proiezione, la trasformazione tra modelli *raster* e vettoriali, ecc.) che alterano la qualità del dato originario;

- Conversione da un formato di archiviazione a un altro che introduce modificazioni difficilmente valutabili;
- Osservazioni imprecise o utilizzo improprio di apparecchiature (ad esempio, l'accuratezza del posizionamento GPS dipende da una corretta scelta del *Datum* e dalla qualità di ricezione del segnale satellitare).

Sono errori che dipendono da violazioni di norme metodologiche e procedurali e che si ripercuotono sulla qualità complessiva dei geodata.

4.4. Geo Big Data

La fase della raccolta di dati secondari ha subito un profondo mutamento nel corso degli ultimi anni [O'Brien *et al.*, 2009]. I dati con cui oggi i ricercatori hanno a che fare sono grezzi e intrusivi, profondamente interrelati tra loro [Montjoye *et al.*, 2012]. I geodata lo sono anche di più poiché le loro componenti si connettono con più facilità a dati di altra natura e coi quali condividono la base geografica di riferimento.

Come spesso ribadito, la rivoluzione geospaziale sta trasformando la ricerca sociale perché ha cambiato il modo in cui la società svolge le sue azioni, il modo in cui viaggia, prende decisioni, condivide le sue storie [Bizer e Schultz, 2010].

A seguito della diffusione dei *social network*, è cambiato radicalmente il modo in cui i soggetti contribuiscono a creare e disseminare contenuti. Spesso, le informazioni prodotte e veicolate da tali strumenti contengono un corredo geospaziale: a un *tweet* può essere associato un set di precise coordinate che permette al ricercatore di conoscere la localizzazione dell'utente e, quindi, quale sia il luogo da cui quei contenuti sono stati prodotti e immessi nel *web*. Gli umani agiscono sempre più come *cyborg sensor* [Vieweg *et al.*, 2010]. Le informazioni che elaborano quando bloggano o postano un contenuto sono ricche di elementi che permettono all'analista di configurare le distribuzioni spaziali dei *social network*, il modo in cui si evolvono nel tempo reagendo agli eventi e adattandosi.

Il *geotagging*, insomma, è una forma di *geocoding* poiché consiste nell'assegnazione di coordinate, catturate da un dispositivo GPS, alla posizione

dei soggetti [Sauer mann e Cyganiak, 2008]. È la manifestazioni tangibile di come le persone agiscano sempre più come sensori che si esprimono in tempo reale. Il contenuto geografico dei *feed* provenienti dai *social media* rappresenta un nuovo tipo di informazione geografica definita AGI (*Ambient Geographic Information*) [Stafanidis *et al.*, 2011]. Estrarla significa poterne analizzare i contenuti, identificare la struttura delle reti sociali, mappare il modo in cui le idee e le informazioni si diffondono nello spazio, delineare l'impronta delle opinioni degli individui in risposta a specifici argomenti o eventi.

I dati riguardanti le città e le persone che vi vivono sono così numerosi e complessi che rimane, tutt'oggi, una sfida, per i ricercatori sociali, cercare di trattare tale complessità in maniera efficiente [Palen *et al.*, 2010; Starbird e Palen, 2011]. Ma se, per lungo tempo, l'analisi geospaziale si è concentrata sulla rilevazione e il trattamento di dati riguardanti le strutture fisiche disposte staticamente sul territorio (edifici, strade, infrastrutture) ignorando i soggetti che in tali aree vivono e si muovono, adesso, gli scienziati dei dati – e in particolare gli scienziati sociali - hanno l'opportunità di osservare i paesaggi umani come organismi viventi verificando come le informazioni si disseminino tra i gruppi, come si dispongono i *cluster* di individui che condividono opinioni e attitudini.

Per analizzare i contenuti geospaziali provenienti dal *web*, è necessario ricondurli dall'amorfo *cyberspace* a un sistema di coordinate univoco.

La forza dei *social media* nel riportare e disseminare tali informazioni è stata dimostrata durante la Primavera Araba, nel 2011 [Pollock, 2011], nel corso della quale *Twitter*, 5 anni dopo il suo lancio, annunciò che più di 200 milioni di individui avevano registrato il proprio *account*, che 100 milioni di questi erano utenti attivi che usufruivano della piattaforma almeno una volta al mese, che 50 milioni di questi la utilizzavano giornalmente. Oggi, nei momenti di particolare criticità (disastri ambientali, terremoti, ecc.), si registra una media di 140 milioni di *tweet* al giorno [Pingdom.com, 2013⁵³].

L'incremento delle comunità *social* e dell'ammontare dei dati che vengono prodotti attraverso tali applicazioni impressiona soprattutto gli analisti. Mentre gli

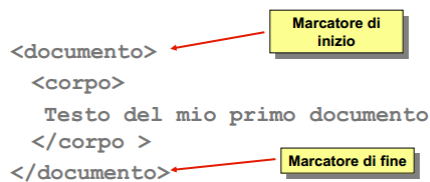
⁵³ <https://www.pingdom.com/>

approcci tradizionali di raccolta e analisi dei dati sono statici poiché viene delineata una specifica strategia che guida la raccolta dei dati necessari, i *Geo Big Data* (in particolar modo quelli prodotti attraverso i *social media*) spingono gli analisti a mettere in discussione molte di quelle che sin ad ora sono state considerate le potenzialità degli strumenti a disposizione della ricerca sociale.

4.4.1. Geolocalizzare i tweet

È possibile rintracciare i *Twitter geodata* nella forma di precise coordinate geografiche associate al *tweet* o di descrizioni prodotte dagli utenti stessi [Eriksson *et al.*, 2010; Poese *et al.*, 2011]. Tuttavia, in alcuni casi, sfruttare tali potenzialità esplicite del *social* non è sufficiente. Secondo gli studi di Stefanidis [2011], la percentuale dei *tweet* geolocalizzati attraverso coordinate o descrizioni esplicite varia dai due terzi alla metà del totale di quelli inviati. Di questi, solo il 16% riporta informazioni a un livello di dettaglio maggiore (ad esempio, descrizioni riportanti il nome della città a cui i *tweet* si riferiscono). Nel caso in cui si abbia necessità di accedere a informazioni più accurate e precise, le tecniche di geolocalizzazione tramite indirizzi IP possono rappresentare la soluzione.

Ad esempio, il *World Wide Web Consortium* (W3C) *Geolocation Application Programming Interface* (API) accede a informazioni provenienti dal *web* in maniera dinamica riportando la localizzazione dalla quale il *tweet* in quel momento è stato inviato. L'estrazione dei *feed* dei *social media* dai *provider* di dati può essere effettuata attraverso le *query* [Beckett, 2004]. L'XML (*Extensible Markup Language*), sviluppato proprio dal *World Wide Web Consortium*, è un linguaggio che consente di creare, memorizzare e diffondere documenti digitali. È un linguaggio a marcatori, composto cioè da istruzioni, definite *tag*, che descrivono la struttura e la forma del documento. Un documento XML è leggibile da un utente umano senza la mediazione di uno specifico software [Berners-Lee *et al.*, 2006].



Una delle tecnologie ideate per il recupero delle informazioni memorizzate all'interno di un file *XML* è *XQuery*. Tale strumento permette di avanzare delle *query* al documento proprio come avviene con il linguaggio SQL, nel caso dei database relazionali. Ad esempio, una delle espressioni principale utilizzata in *XQuery* per formulare interrogazioni complesse e che viene chiamata espressione FLWR (*For-Let-Where-Return*) è una generalizzazione del costrutto SELECT FROM (WHERE) del linguaggio SQL [*Ibidem*].

In risposta alle *query* possiamo ottenere dal server *metadata* o associazioni di *metadata* e dati. Per quanto riguarda *Twitter*, ad esempio, i dati offerti dal *server*, e ricevibili in risposta ad una *query*, sono veri e propri *tweet* e *metadata* associati (informazioni sull'utente, orario e data di pubblicazione del *tweet*, luogo dal quale è stato inviato in rete). La risposta alla *query* può fornirci anche informazioni aggiuntive sulla natura del *tweet*: se si tratta, cioè, di un *retweet* o di una risposta ad un altro *tweet*.

```

{
  "retweet_count": 3,
  "retweeted": false,
  "geo": {
    "type": "Point",
    "coordinates": [55.76515269, 37.59717267]
  },
  "possibly_sensitive": false,
  "coordinates": {
    "type": "Point",
    "coordinates": [37.59717267, 55.76515269]
  },
  "location": {
    "name": "URB11URB31URB41URB31URB12URB10",
    "url": "https://api.twitter.com/1.1/geo/2014/38202113698271360",
    "attributes": {},
    "full_name": "URB11URB31URB41URB31URB12URB10",
    "country": "URB42URB2URB41URB41URB31URB41",
    "country_code": "RO",
    "place_type": "city",
    "bounding_box": {
      "type": "Polygon",
      "coordinates": [[[37.1216733, 55.4988341], [37.5659116, 55.4988341], [37.9606116, 55.8291932], [37.1216733, 56.8291932]]]
    }
  },
  "id": "43334343434343"
}

```

A red box highlights the "geo" object in the JSON:

```

"geo": {
  "type": "Point",
  "coordinates": [55.76515269, 37.59717267]
},

```

(Figura 4.9. Estratto di metadata di un tweet, ottenuto da un API twitter, che riporta le coordinate dell'utente)

L'estrazione delle informazioni dai *feed* dei *social media* prevede [Tomaszewski e MacEachren, 2010]:

- 1- L'estrazione dei dati dai *provider* (i diversi *server social media*) attraverso le API;
- 2- L'inserimento dei dati nel *database* dedicato;
- 3- L'analisi dei dati ed estrazione di quelli di interesse del ricercatore.

Considerata la complessità di tali strumenti, l'analista potrà decidere di utilizzare i numerosi strumenti online di *Social Media Management* (ad esempio, <https://hootsuite.com/>) che non permettono, però, di lavorare su grandi database.

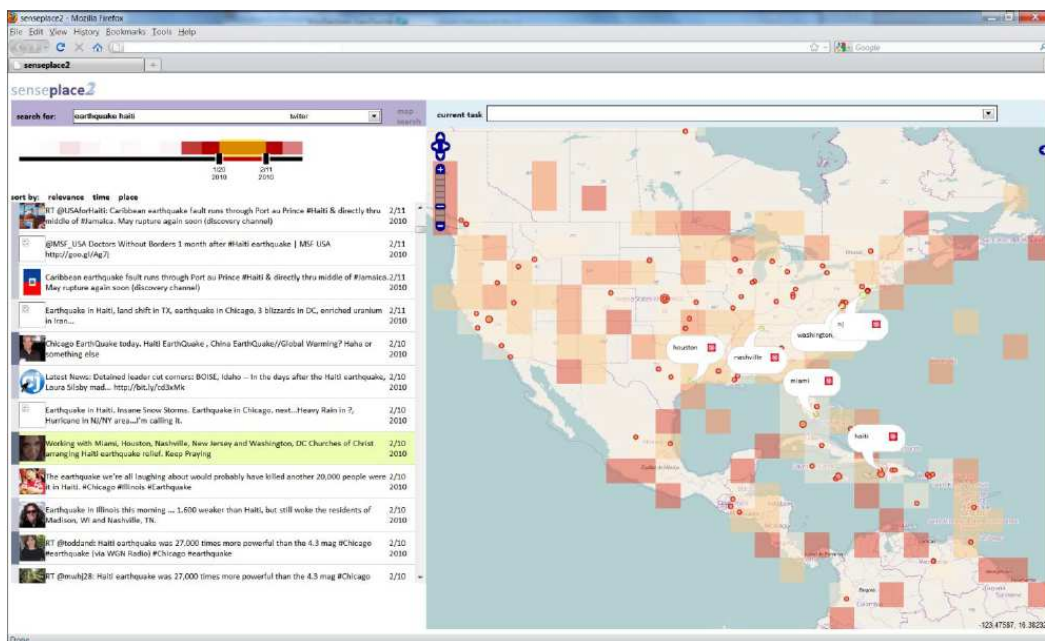
Esistono numerose *web application* create appositamente per mappare i *tweet*: *Compepi*⁵⁴ - *Computational Epidemiology Research –Neoformix*⁵⁵ o il prototipo *SensePlace2* sono solo alcuni esempi [Sakaki *et al.*, 2010; Mendoza *et al.*, 2010]. In particolare, *SensePlace2* permette di sviluppare *twitter map* dettagliate filtrando per *hashtag*. L'applicazione funziona in modo da avanzare ogni giorno al server *twitter* le *query*, in base a una serie di termini chiave che i ricercatori scelgono come rilevanti per la loro indagine. In risposta, riceveranno i *tweet* e i *metadata* ad essi associati che verranno archiviati all'interno di un database SQL. Le entità archiviate vengono separate in apposite tabelle. Un'applicazione collegata al database produce la rappresentazione cartografica dei *tweet* [Berners-Lee *et al.*, 2006].

Nonostante le potenzialità di tali strumenti, spesso non è possibile accedervi o utilizzarli liberamente. Il caso di *SensePlace2* è emblematico poiché i risultati della ricerca della *PennState University* e *GeoVista* e finanziata dall'*Engineer Research and Development Center*⁵⁶, non sono ancora accessibili. Tantomeno, è possibile utilizzare l'applicazione per i propri scopi di ricerca [Tomaszewski, 2008].

⁵⁴ <http://compepi.cs.uiowa.edu/>

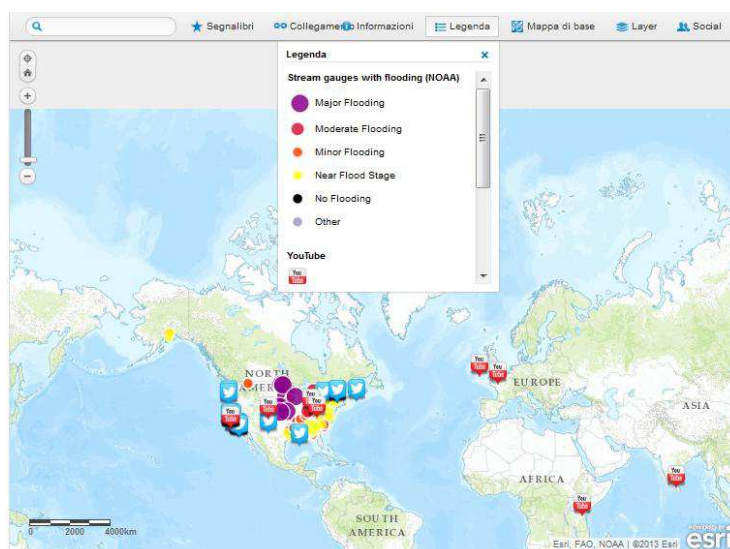
⁵⁵ <http://www.neoformix.com/>

⁵⁶ <http://www.erd.usace.army.mil/>



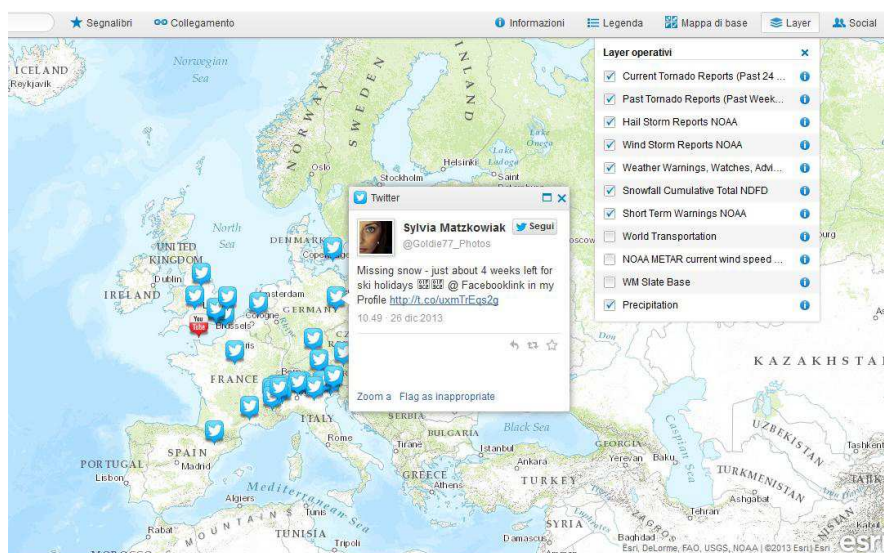
Molti dei *geo twitter data* prodotti in tempo reale sono accessibili invece attraverso ArcGIS online. Cliccando sull'indirizzo: <http://www.esri.com/services/disaster-response/severe-weather/latest-news-map>, si avrà, ad esempio, la possibilità di accedere ai dati riguardanti i fenomeni naturali (precipitazioni, alluvioni, tempeste di neve, terremoti, ecc.). Cliccando sul comando *Informazioni*, che si trova nella barra in alto, si avrà accesso ai *metadata*. Si potrà scoprire, così, chi li ha prodotti⁵⁷, come e quando. Strati di informazioni possono essere aggiunti o sottratti cambiando scala, colori, simboli e classificazione dei dati. Non sono semplici mappe di riferimento in grado di rispondere a domande come: «Dove ci troviamo in questo momento?» ma possono essere utilizzate come strumenti di indagine geografica, necessari a comprendere i modelli spaziali e temporali dei fenomeni in esame [Berners-Lee *et al.*, 2006].

⁵⁷ Tali dati sono prodotti principalmente dal NOAA's National Weather Service Storm Prediction Center (USA) e dall'USGS (U.S. Geological Survey).



(Figura 4.6. Snapshot ArcGIS online)

Cliccando sul comando *Social*, posto in alto a destra, è possibile scegliere i *social network* (*Flickr*, *Twitter*, *YouTube*) dai quali vengono prodotti i dati relativi ai fenomeni rappresentati sulla mappa. Cliccando sulla relativa icona a forma di ingranaggio, posta accanto al *social network*, si potrà digitare il *tag* (o *hashtag*) di cui interessa conoscere la distribuzione nello spazio.



(Figura 4.7. Snapshot ArcGIS online)

4.5. Raw OpenGeoData Now!

In Italia, è rilevante l'esperienza del gruppo di professionisti che ha fondato il portale *DatiOpen.it*⁵⁸. L'obiettivo di tale iniziativa è quello di raccogliere e documentare la maggiore quantità possibile di dati *open* italiani. La redazione opera quotidianamente allo scopo di monitorare la produzione di tali dati e di identificarne di nuovi da catalogare e da inserire nel sistema. I dati disponibili sono scaricabili e visualizzabili in tabelle, grafici e mappe interattive.

Oltre a *DatiOpen*, si fanno sempre più numerose le associazioni che seguono, come obiettivo primario, l'apertura dei dati geografici prodotti dalla Pubblica Amministrazione. *OpenGeoData*⁵⁹, che svolge funzione di controllo e invita gli Enti Locali a liberalizzare i dati, *GEOforUS*,⁶⁰ *e-Geo*,⁶¹ sono solo alcuni esempi di portali ai quali i ricercatori sociali possono accedere per recuperare dati riguardanti l'ambiente, l'arte e la cultura, il commercio, la demografia, l'economia, le infrastrutture e i trasporti, l'istruzione, il lavoro, la politica, la sanità, il sociale, lo sport e il tempo libero, il turismo.

Le iniziative associative, che si focalizzano sul coinvolgimento di comunità di residenti ed esperti GIS nel processo di mappatura del territorio, sono spesso sostenute dal governo centrale che comincia a gestire i processi decisionali tenendo conto del loro contributo.⁶² I volontari mappano i territori in cui vivono, localizzano gli oggetti che vi si trovano utilizzando GPS, digitalizzando strade e *feature* [Zook *et al.*, 2010]. Ad aver favorito questo processo di creazione e condivisione di geodata, la tecnologia di localizzazione di tipo *consumer* [Estes e Mooneyhan, 1994], ora ampiamente disponibile. Ognuno di noi crea nuovi dati

⁵⁸ <http://www.datiopen.it>

⁵⁹ www.opengeodata.it

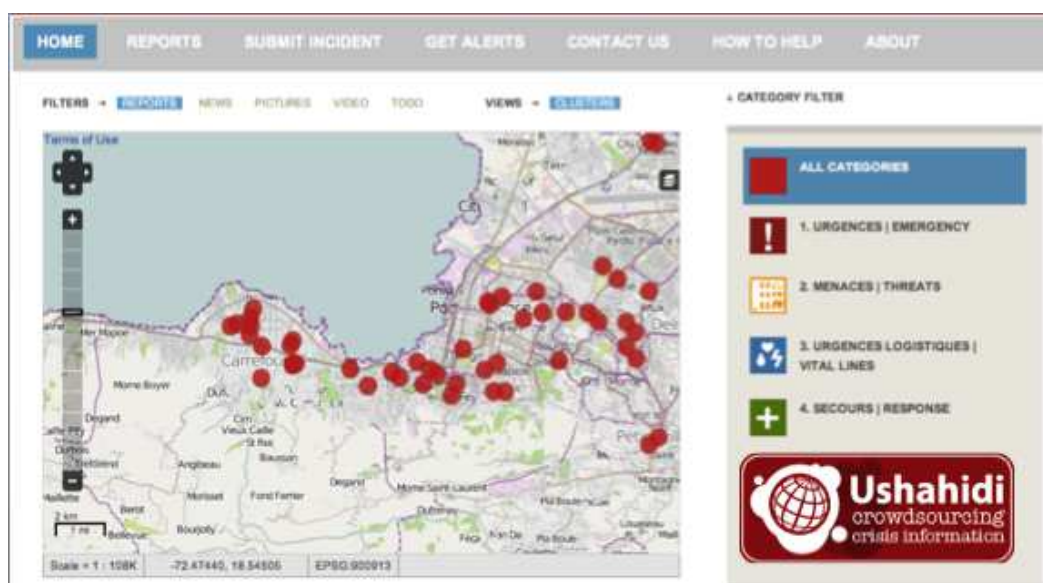
⁶⁰ <http://www.geoforus.it> mette a disposizione dati stradali di 50 città italiane in formato shape e complete di toponomastica. I grafi possono liberamente utilizzati senza vincolo.

⁶¹ <http://www.egeo.unisi.it/>, il primo e più completo motore di ricerca online di carte geomantiche italiane.

⁶² Un esempio importante è dato dall'esperienza OpenPompei, il progetto finanziato dal PON Governance e Assistenza Tecnica 2007-2013, nell'ambito di una collaborazione con il Ministero dello Sviluppo Economico. Per le caratteristiche fortemente sperimentali del progetto e la sua forte connotazione in tema di *civic hacking* ed apertura dei dati della Pubblica Amministrazione, i referenti del progetto hanno deciso di aprirsi ad una comunità di esperti e di chiedere un loro contributo.

territoriali ogni giorno [Brown, 2006]. Perciò, oltre ai governi e alle industrie, sino a qualche decennio fa principali produttori di dati territoriali difficilmente accessibili, si affaccia una nuova comunità di produttori e utilizzatori di dati aperti. Tra questi, anche i ricercatori cominciano a dare il loro contributo nel processo di elaborazione della cosiddetta VGI (*Volunteer Geographic Information*) [Starbird e Palen, 2011; Butler, 2006].

Durante il terremoto di Haiti del 2010, i *mapper* hanno utilizzato un sistema di produzione di dati cartografici, chiamato *Ushahidi*⁶³, oggi molto diffuso tra gli attivisti sociali, *citizen journalist* e esperti GIS. Questa esperienza cambiò definitivamente la conformazione di una *crisis map*.



(Figura 4.2. Ushahidi 24 ore dopo il terremoto di Haiti)

Patrick Meier, un *PhD student* della Tufts University di Boston, utilizzò *Hushahidi* caricando, sulla piattaforma GIS disponibile sul sistema, le informazioni georiferite riportate nei *tweet* inviati da chi stava vivendo, in quei momenti, la tragedia. A Patrick, il giorno dopo, si unirono molti altri volontari ricercatori e un centinaio di studenti della Tufts, i quali vennero presto definiti i primi *digital humanitarians* [Goodchild *et al.*, 2007].

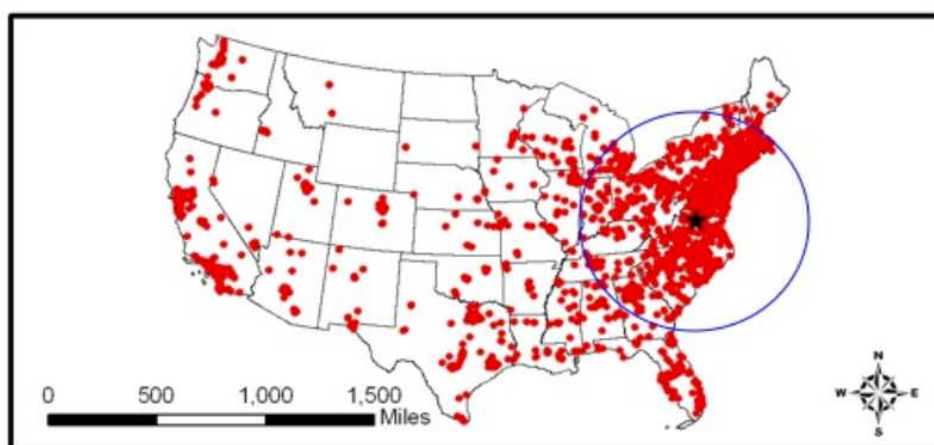
⁶³ <http://www.ushahidi.com/>



(Figura 4.3. Un tweet che comunica la riapertura di una farmacia durante il terremoto di Haiti. Credits. Twitter)

Stefanidis *et al.* [2011] raccolsero i *tweet* elaborati dagli utenti nei momenti immediatamente successivi al terremoto di magnitudo 5.8, avvenuto in Virginia, il 23 Agosto 2011. La Figura 4.4. mostra i punti geografici da cui sono stati elaborati e immessi sulla rete i *tweet* riferiti al terremoto, nella prima ora successiva all'accaduto.

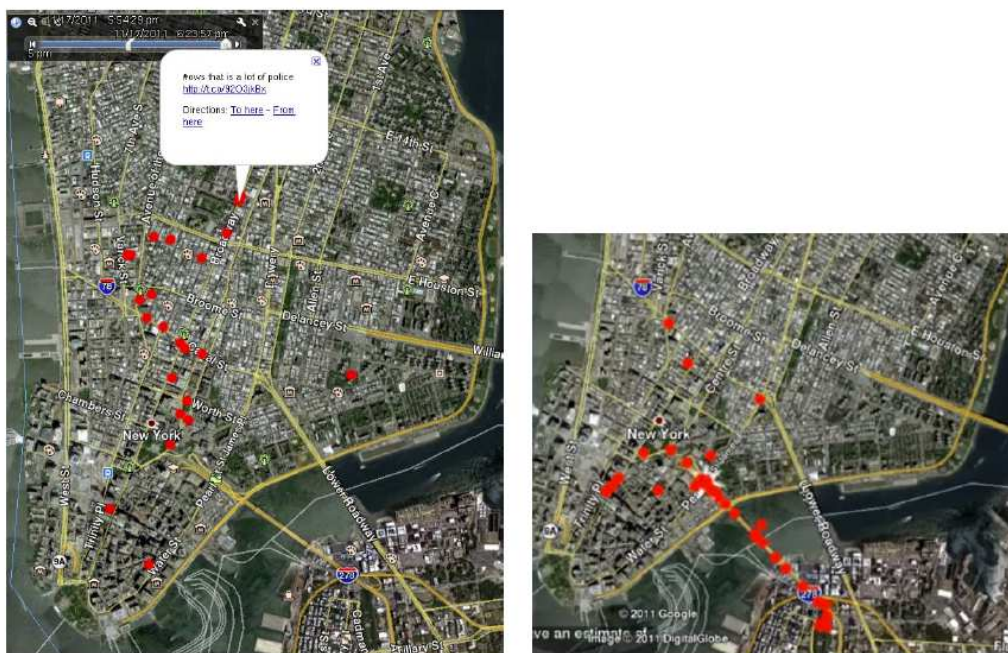
Tweets Within 60 Minutes After Earthquake



(Figura 4.4. Localizzazione tweet nei 60 minuti dopo il terremoto in Virginia)

La stella identifica l'epicentro del terremoto; il cerchio blu indica l'area nella quale è stato percepito dalla popolazione.

I *Social Media* possono aiutare a monitorare gli eventi, come grandi manifestazioni o proteste [Zook *et al.*, 2010]. La Figura 4.5. mostra i punti della città di New York dai quali sono stati elaborati e immessi in rete i *tweet* riferiti alla manifestazione *Occupy Wall Street* del 17 Novembre 2011. I *tweet* descrivono chiaramente due momenti della marcia lungo il Ponte di Brooklin. L'immagine a sinistra riporta i *tweet* del pomeriggio, momento in cui i manifestanti stanno per spostarsi verso il ponte; quella a destra riporta quelli elaborati a tarda sera.



(Figura 4.5. Localizzazione tweet durante Occupy Wall Street)

Che i dati geografici attirino l'attenzione degli analisti è testimoniato dal numero di *download* effettuati dal portale <http://www.data.gov>, voluto fortemente da Barack Obama, sin dal suo primo mandato [OpenGeoData, 2013]. È più che chiaro che le informazioni geografiche facilitino i processi decisionali e, specie se le comunità sono coinvolte nei processi partecipativi di mappatura del territorio, migliorano la qualità della vita dei soggetti che vi vivono. Occorre però che, oltre ad essere disponibili, i dati siano pubblicati in formato *open*.

Perché un *dataset* sia considerato *open* è necessario che sia [Alvisi *et al.*, 2013]:

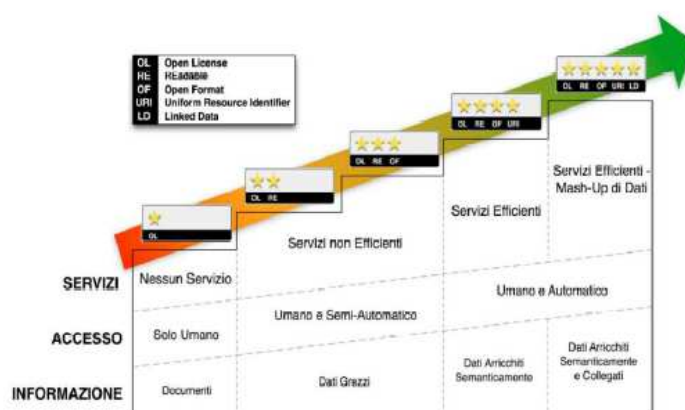
- Completo. I dati devono essere resi disponibili nella loro interezza e comprendere i metadati. I formati devono poter essere esportabili, utilizzabili e aggregabili;
- Elementare. I dati devono essere pubblicati rispettando la loro caratteristica di granularità (*raw data*). Devono essere pubblicati così come vengono raccolti, evitando eventuali rielaborazioni;
- Aggiornato. I dati devono essere disponibili con tempestività e continuità preservandone il valore e l'utilizzo;
- Accessibile. I dati devono poter essere disponibili a un ampio numero di utenti;
- *Machine-readable*. I dati devono poter essere elaborati in modo automatizzato;
- Non discriminanti. I dati devono essere accessibili senza necessità di registrazione dell'utente nel sito che li raccoglie;
- Non-proprietari. Se i dati sono in formato proprietario, l'utente dovrà poter accedere all'apposito software in grado di leggere il dataset.
- Riutilizzabili. I dati devono essere liberi da *copyright*, licenze o marchi e non essere sottoposti a segreto commerciale;
- Permanenti. Le caratteristiche elencate devono persistere per tutto il ciclo di vita del dataset.

Il livello di *openness* è stato meglio classificato nel 2007 dall'informatico britannico Tim Bernes-Lee nel suo blog *Open Knowledge Foundation*⁶⁴. I ricercatori accedono a dati a una stella, se questi sono leggibili e stampabili, con licenza d'uso. I dati di tale natura non sono strutturati e non è possibile leggerli o rielaborarli in modo automatico. In genere, si tratta di dati in formato *.pdf* o *.jpg*⁶⁵. I dati a due stelle, invece, sono strutturati ma in formato proprietario. Possono

⁶⁴ <http://blog.okfn.org/2007/11/07/give-us-the-data-raw-and-give-it-to-us-now/>

⁶⁵ Sono molto comuni i casi in cui l'analista GIS è costretto a lavorare su file cartografici in formato *.pdf*. Si tratta di dati resi disponibili dai diversi enti locali, restii a pubblicarli in formato aperto. Per capire come masterizzare e georeferenziare una base cartografica in formato pdf, basta fare riferimento all'ultimo paragrafo del secondo capitolo della presente tesi.

essere rielaborati, quindi, a patto che si disponga del software adatto a gestirli⁶⁶. I dati a tre stelle sono riutilizzabili, strutturati e non proprietari. I dati a quattro stelle sono strutturati e codificati in un formato non proprietario (ad esempio, .csv⁶⁷) e sono individuabili e utilizzabili online. Chiunque può utilizzare dataset pubblicati e ai quali è stata associata una licenza d'uso.⁶⁸ I dati a cinque stelle sono i cosiddetti *Linked Open Data*; rendono, cioè, possibile il collegamento ad altri dataset esterni di natura eterogenea [Tim Berners-Lee *et al.*, 2007; Bizer, 2010]. Oltre ad essere *Machine-Readable*, sono indicizzabili sui motori di ricerca e inseriti nei portali di dati aperti.



(Figura 4.8. Classificazione Openness dei dati)

Dopo aver pubblicato i dataset aperti è necessario, infatti, diffonderli non solo sui motori di ricerca ma anche sui siti specializzati. Alcuni Enti e associazioni stanno lavorando all'unificazione dei cataloghi dei dati a loro disposizione utilizzando un'unica terminologia per denominare i modelli *standard* dei dati.

È difficile che tutti i dataset di interesse del ricercatore provengano dalla stessa fonte o che siano stati raccolti nello stesso intervallo di tempo o in base alla

⁶⁶ Poiché i *dataset* pubblicati su Internet possono nel tempo essere aggiornati, modificati, associati a nuove licenze, è auspicabile ad essi siano associati i Metadati.

⁶⁷ Il .csv è un formato dati basato sui file di testo per lo scambio di una tabella dati.

⁶⁸ Per la riproduzione di geodata aperti, le licenze utilizzate più comunemente sono la *Creative Commons 0 (CC0)*, la licenza *open* per eccellenza, la *Creative Commons Attribuzione (CC-BY)*, per la quale bisogna dichiarare la paternità dell'opera, e l'*Italian Open Data License 2.0 (IODL 2.0)*, simile alla CC-BY ma disponibile solo in lingua italiana.

stessa scala. Nonostante questa imperfezioni di fondo, le mappe sono ugualmente utili. I dati che il ricercatore sociale riuscirà a raccogliere saranno comunque utili [Field e O'Brien, 2010]. Occorrerà solo che questi si ponga le giuste domande e che sia consumatore e produttore critico di dati aperti, che medi tra la necessità di accedervi e lo sforzo richiesto per trattarli e ristrutturarli logicamente e concettualmente.

CAPITOLO QUINTO CAQDAS E QUALITATIVE GIS

I primi tentativi di integrazione tra dati qualitativi e GIS risalgono all'inizio dello scorso decennio, quando ricercatori di diverse discipline, antropologi, psicologi e geografi umani, cominciarono a riflettere sulle potenzialità di questo connubio contribuendo alla messa a punto di quello che sarebbe presto diventato il *Qualitative GIS Approach* [Al-Kodmany, 2002; Shiffer, 2002; Weiner e Harris, 2003].

Qualche anno dopo, anche gli studiosi orientati a una ricerca sociale di tipo qualitativo e interpretativo cominciarono ad analizzare congiuntamente dati eterogenei (video, mappe concettuali, tracce audio, stralci di interviste, foto, waypoint GPS⁶⁹, ecc.) utilizzando, in maniera non convenzionale, sia software specifici per la ricerca qualitativa che strumenti GIS [Elwood, 2006, 323-341; Crang, 2005, 225-233].

Ma il *Qualitative GIS Approach* è molto più che una semplice integrazione tra software. I ricercatori che se ne avvalgono assumono che l'interpretazione dei dati utili all'analisi venga prodotta e negoziata seguendo un processo non lineare, ma iterativo e circolare [Seidel, 1998, 2], soggetto a continue modificazioni. Come evidenziato nei capitoli precedenti, il ricercatore che si avvale esclusivamente dei software GIS procede attraverso una sequenza di fasi standardizzate, entro una

⁶⁹ Il GPS serve a localizzare le posizioni e misurare le distanze. Il sistema è stato progettato dagli Stati Uniti, a partire dal 1970, originariamente per scopi militari. Appartiene alla grande famiglia dei sistemi globali di navigazione satellitare (GNSS) di cui fanno parte il russo GLONASS e il Galileo dell'Unione Europea. Ogni GNSS funziona in base allo stesso principio: una rete di satelliti nello spazio trasmette segnali sulla terra che veicolano informazioni sulla posizione, in base alla posizione del satellite, e al momento in cui il segnale è stato inviato. Oggi, è possibile combinare questi segnali con quelli *wi-fi* e da altre fonti per migliorare la precisione e la copertura. La diffusione di strumenti Virtual Globe, come Google Earth, ha reso l'accessibilità alle immagini della terra più semplice di prima. Un satellite nello spazio può registrare dati riguardanti gli oggetti visibili alla luce e quelli invisibili dello spettro elettromagnetico (attraverso infrarossi, per esempio) e assegnare valori ad ogni cella della griglia per sviluppare un'immagine. La scienza e la tecnologia associata all'*imaging* della terra dall'alto si chiama *Remote Sensing*. È una nuova disciplina che si focalizza sullo sviluppo di nuovi modi di rappresentare la terra, di interpretare e analizzare le immagini stesse.

cornice lineare [Dainelli *et. al.*, 2010]. La relazione tra le fasi che compongono il processo di ricerca è di tipo unidirezionale. Il punto di partenza dell'analisi è costituito da un'ipotesi da verificare, a partire da un impianto teorico pre-esistente, attraverso un'attività di confronto con i dati oggettivi. I GIS sono usati comunemente per analizzare e rappresentare dati territoriali, i quali difficilmente permettono di esplorare le esperienze vissute dai residenti o il loro livello di *place attachment* [Kasarda e Janowitz, 1974]. Il ricercatore che intende condurre un'analisi GIS di tipo standard, si limiterà a pensare alla superficie terrestre non come un luogo della performance dei fenomeni sociali ma come uno spazio fisico dotato esclusivamente di alcune proprietà oggettive [Dainelli *et al.*, 2010, 96]. Lo spazio fisico sarà ricondotto all'interno di un modello che permetterà l'analisi e la gestione dei dati e nel quale occorrerà distinguere nettamente i fenomeni dagli elementi geografici.

Seguendo il *Qualitative GIS Approach*, invece, il ricercatore potrà più volte tornare sui suoi passi, dalla fase di raccolta dati, e di una concomitante interpretazione degli stessi, fino al raggiungimento della saturazione teorica attraverso la produzione di teorie concettualmente dense [Cicognani, 2002]. Il processo di codifica *preliminare*, che caratterizza la GIS analysis, verrà continuamente messo in discussione dalla codifica *a posteriori*, tipica della ricerca qualitativa. Il ricercatore, da un lato, avrà la possibilità di riprodurre preliminarmente la realtà sociale sistematizzando in un ambiente GIS i dati che ad essa si riferiscono; dall'altro, grazie alle comparazioni cicliche che caratterizzano il processo di analisi qualitativo, potrà occuparsi di riassemblare tali dati e di produrre il modello teorico-interpretativo. La stessa definizione di approccio qualitativo ai dati, del resto, è di tipo aperto perché, pur prevedendo la necessità di interpretarli, non esclude che anche nel processo di codifica si faccia riferimento al dato quantitativo. «Un approccio qualitativo è un approccio in cui vi è la necessità di interpretare i dati attraverso la identificazione e la possibile codifica di temi, concetti, processi, contesti, ecc., al fine di costruire spiegazioni o teorie, o di testare o ampliare una teoria» [Lewins e Silver, 2004, 2 in Trobia, 2005, 116]. In egual modo, le regole di un'analisi GIS di tipo standard, si sposano bene con

quelle previste dall'analisi qualitativa computer-assistita dato che anche il ricercatore che si avvale dei CAQDAS dovrà formulare delle ipotesi iniziali, seppur soggette a revisioni continue.

Il *Qualitative GIS approach* lascia emergere, ancora una volta, la riduttività di una visione dicotomica che definisce i modelli qualitativo e quantitativo come totalmente separati e opposti [Kruglanski e Jost, 2000, 45-67]. Collocandosi nel cosiddetto *continuum* qualità-quantità [Denzin, 1978], a difesa del pluralismo metodologico [Mannetti, 1998], facilita l'applicazione di strategie che spingono il ricercatore verso continui aggiustamenti, nei diversi momenti della ricerca. Supporta le sue capacità creative e interpretative, sin dall'inizio del processo di analisi, in cui avviene la sistematica organizzazione dei dati raccolti; facilita l'esplicitazione di procedure rigorose, indispensabili sia all'ispezionabilità della base empirica [Ricolfi, 1997; Mazzara, 2002, 21-41], che all'utilizzo di entrambi i tipi di software.

Ma il vantaggio più rilevante del *Qualitative GIS Approach* sta nella possibilità di condividere e confrontare il lavoro tra ricercatori diversi che usano software diversi, nella fattispecie CAQDAS e GIS, purché seguano le stesse procedure in maniera sistematica. Ciò, a conforto del rigore metodologico che, nelle scienze sociali, si manifesta soprattutto grazie alla collaborazione di gruppo e all'interdisciplinarietà. L'integrazione QDA-GIS, in particolare, costringe sia all'accordo tra diversi ricercatori sui significati da assegnare alle categorie, sia al confronto tra diversi sistemi di codifica.

5.1. CAQDAS

I *Computer Assisted Qualitative Data Analysis Software* (CAQDAS) [Fielding e Lee, 1998] consentono al ricercatore di gestire i dati qualitativi raccolti, semplificando l'intero processo di analisi, e di lavorare su grandi quantità di informazioni che vengono sistematizzate, conservando la loro ricchezza e complessità. Attraverso l'utilizzo delle numerose funzioni e procedure software di manipolazione dei dati, il ricercatore è in grado di dedicarsi al lavoro di interpretazione degli stessi [Silverman, 2002], risparmiando tempo e fatica.

Nel corso dell'ultimo ventennio, i pacchetti CAQDAS che supportano la gestione dei dati qualitativi sono diventati sempre più numerosi. Ne esistono di diversi tipi tra *word processor* (che svolgono semplici funzioni di video scrittura), *text retriever* (che ricercano informazioni restituendo i testi in cui esse si trovano ed effettuando un conteggio delle occorrenze), *code-and-retriever* (che codificano le informazioni contenute in testi di natura diversa quali immagini, video, stralci di interviste, ecc.), *code-based theory builder* e *conceptual network builder* (che, oltre a possedere le funzioni sopra elencate, permettono di creare connessioni tra i temi o i codici e di visualizzare tali connessioni. Inoltre, sono dotati di funzioni e strumenti di elaborazione diagrammi e mappe semantiche) [Roebuck, 2013]. Tali software, insomma, supportano il ricercatore durante le fasi di registrazione e archiviazione dati, codifica e attribuzione di un'etichetta di riconoscimento alle informazioni che vengono isolate dal resto del testo. Permettono, inoltre, di recuperare, analizzare i dati e visualizzare i risultati emersi dall'analisi, attraverso rappresentazioni grafiche che riproducono le connessioni logiche tra i temi.

La struttura organizzata, nella quale vengono raccolti i *corpora* testuali, permette al ricercatore di riflettere, concettualmente e teoricamente, giocando sull'interscambio reciproco tra teoria e dati, riorganizzando il materiale in base alle esigenze di ricerca e conducendo, in definitiva, un lavoro rigoroso di analisi. Occorre, però, sottolineare che «software cannot pull good work out of a poor qualitative researcher, but it can in fact help competent researchers do more consistent and thorough research» [Weitzman, 2000, 817]. Come suggerisce l'autrice, tali software aiutano il ricercatore a sviluppare proposizioni concettuali ma non automatizzano l'analisi qualitativa.

Il meccanismo di interscambio continuo e reciproco tra teoria e dati, prima descritto, è lo stesso sul quale si basa la *Grounded Theory*⁷⁰ [Strauss e Corbin, 1990; Strati, 1997, in Ricolfi 1997, 125-63]: il metodo del confronto costante permette di generare proposizioni ipotetiche a partire dai dati qualitativi

⁷⁰ Secondo Glaser e Strauss [1967], le teorie vanno ricercate all'interno dei dati in cui sono radicate, ancorate (*grounded*, appunto). Il ricercatore è privo di teorie pre-esistenti o conoscenze che possano influenzare il processo di scoperta. Sono i dati a far emergere un modello teorico. Il ricercatore qualitativo cerca le sue teorie implicite nei dati raccolti concentrandosi maggiormente sulla fase di codifica degli stessi.

[Charmaz, 1995, 27-49]. Il ricercatore non testa ipotesi da corroborare ma mette in atto processi di analisi iterativa e induttiva che permettono alla teoria di emergere dai dati [Strauss e Corbin, 1998, 12-14]. È una tecnica che si concentra sull'esplorazione delle possibilità multiple e sull'interpretazione. «Per il ricercatore, non esistono dati da estrarre dalle persone, come diamanti sepolti nelle profondità di una miniera, da portare alla luce il più possibile puri, integri e intatti dagli interessi del ricercatore stesso, dal soggetto e dall'ambiente sociale e culturale in cui la ricerca si muove» [Mantovani, 2003, 26].

Il cuore della *Grounded Theory* è il processo di codifica [Lonkila, 1995] che permette di capire e organizzare, in modo sistematico, i significati insiti nei testi. I codici assegnati dal ricercatore sono le etichette che permettono di associare a un particolare artefatto del dato qualitativo una specifica unità di significato. Il processo di creazione e assegnazione dei codici è diretto a trovare il significato e le relazioni insite nei dati stessi differenziandoli o combinandoli diversamente [Miles e Huberman, 1994].

5.2. QDA-GIS analysis

L'utilizzo dei *Geographic Information Systems*, nell'ambito dell'analisi qualitativa computer-assistita, si inserisce nella strategia "semplificare per capire". La CAQ-GIS *analysis* si basa sull'archiviazione di dati di natura eterogenea, nel database ibrido-relazionale GIS [Jung, 2009, 115-132] o nell'Unità Ermeneutica CAQDAS [Fielding, 2012], facilitando l'analisi induttiva e interpretativa. Ciò, grazie alla possibilità, offerta dai GIS e dai CAQDAS di archiviare, organizzare, analizzare e rappresentare dati qualitativi e quantitativi georiferibili. GIS e CAQDAS, insieme, supportano il ricercatore nel corso della *Qualitative GIS analysis*.

A differenza di quanto affermato dai sostenitori della *Grounded Theory* [Glaser e Strauss, 1967; Strauss e Corbin, 1990], per i quali il ricercatore dovrebbe accostarsi alle unità di analisi senza alcun modello teorico di riferimento, l'analista CAQ-GIS non sarà completamente scevro da qualsiasi prospettiva pre-esistente. L'analisi GIS prevede infatti che, attraverso uno

specifico procedimento logico, composto da concrete operazioni di ricerca, [Boudon e Lazarsfeld, 1969, 41-52], le variabili possano essere costruite anche a partire da un concetto ad elevato livello di astrazione [Marradi, 1987]. Tuttavia, la natura della CAQ-GIS analysis condurrà, comunque, il ricercatore a intraprendere un percorso iterativo tra teoria, dati e interpretazione. I suoi schemi concettuali assumeranno una forma organizzata che potrebbe essere messa in discussione nel corso dell'intero processo di analisi e, soprattutto, nella fase dedicata all'interpretazione.

5.3. Testi e contesti

Individuare le tracce di una qualche forma di *spazialità*, nel dato qualitativo, e di *qualitativeness* [Jung, 2009, 126], nel dato spaziale, è il primo passo da compiere per condurre una CAQ-GIS analysis sistematica. La distribuzione e frequenza nel testo di segmenti portatori di informazioni localizzabili geograficamente permetterà al ricercatore di associare le narrazioni ai luoghi. Rimandiamo ad altri testi per l'approfondimento delle diverse tecniche di analisi statistica e testuale solitamente utilizzate per estrarre dal *corpus* insiemi di parole o sequenze delle stesse [Della Ratta-Rinaldi, 2009 in Cannavò, Frudà, 2009, 141-151]. Qui, è interessante sottolineare che un'analisi CAQ-GIS non può prescindere dall'importanza delle informazioni categoriali geografiche associate ai dati qualitativi testuali poiché risorse indispensabili alla identificazione di codici utilizzabili in entrambi i tipi di software.

Per far sì che la feconda funzione di relazione tra *testi e contesti spaziali* non vada perduta, è necessario che gli analisti coinvolti nel progetto di ricerca esplicitino regole di classificazione chiare che possano essere seguite nelle diverse fasi di un percorso così composito. Tali regole dovranno facilitare lo sviluppo di una spiegazione del fenomeno che renda conto delle relazioni tra i dati e dei processi che organizzano tali relazioni focalizzando l'attenzione sulla dimensione narrativa e spaziale. In particolare, per ciò che concerne l'analisi della dimensione narrativa, essa è un valido strumento in grado di organizzare la struttura percettiva di un luogo così come vissuto da una comunità.

Nello studio delle regole attraverso le quali si struttura una narrazione, Labow [1997, 395-415 in De Gregorio e Mosiello, 2004, 48-49] ha distinto le diverse fasi della struttura narrativa in *Abstract*, *Orientation*, *Complication*, *Evaluation*, *Result*, *Coda*. Nella prima, il narratore/intervistato introduce l'argomento; nella seconda, fornisce dettagli specifici sulle variabili contestuali (luogo, tempo, situazione). La terza (*Complication*) è quella che spesso costituisce il nucleo della narrazione; la quarta (*Evaluation*) consta di una serie di dispositivi valutativi che illustrano l'orientamento del narratore nei confronti degli eventi narrati. Infine, le fasi *Result* e *Coda* rappresentano, rispettivamente, la risoluzione dell'evento e la chiusura della narrazione. Le narrazioni portano con sé tracce spaziali socialmente significative. Pertanto, soprattutto in sede di rilevazione dati, in cui affiorano i contenuti che si esprimono nella fase della struttura narrativa chiamata *Orientation*, è opportuno che l'intervistatore faccia in modo che emergano, in maniera più o meno esplicita, contenuti rilevanti situabili spazialmente, inducendo i rispondenti a descrivere, anche dettagliatamente, condizioni, relazioni e processi che si esprimono nello spazio urbano. In tal modo, si potrà scegliere tra gli elementi testuali che più agevolano la comprensione dell'ordine spaziale della narrazione.

5.4. Qualitativeness

La principale difficoltà nella quale si può incorrere nel corso della CAQ-GIS analysis risiede nella natura eterogenea dei dati da trattare. Testi di natura diversa, tra immagini, tracce video, audio, numeri, mappe concettuali, materiale cartografico, ecc., dovranno essere trattati congiuntamente.

Occorre, inoltre, saper distinguere i dati portatori della cosiddetta *qualitativeness* [Jung, 2009]. I sociologi che, per primi, hanno sperimentato la CAQ-GIS analysis, nell'ambito della ricerca sociale [Gilbert e Masucci, 2005, 277-279; Kwan e Knigge, 2006, 1999-2002], dimostrano che i dati qualitativi, utili a tale tipo di ricerca, sono molto più che semplici dati non numerici. Più dei cartografi, i geografi e gli antropologi sono stati in grado di *leggere* nei dati geografici le impronte del sapere esperienziale dei soggetti e di interpretare i

significati da questi attribuiti alle rappresentazioni dei luoghi abitati. Per Crampton e Krygier [2006], anche il significato di una mappa è socialmente costruito e viene continuamente ri-negoziato da coloro che conoscono ciò che rappresenta. Ogni immagine cartografica, quindi, può essere portatrice di *qualitativeness* se in essa vengono rintracciate informazioni di tipo qualitativo [Fiedler *et al.*, 2006, 145-171].

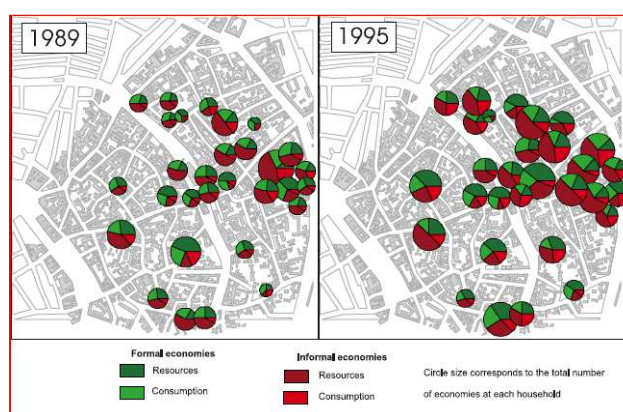
È chiaro che non tutte le immagini cartografiche sono portatrici di informazioni qualitative. Una mappa satellitare o una foto aerea rappresentano il mondo reale riportando gli oggetti che lo caratterizzano geograficamente (strade, edifici, fiumi, parchi, ecc.). Immagini simili non permettono all'analista di ricostruire un quadro completo che ritragga le dinamiche delle relazioni sociali che vengono intrattenute in quello spazio, i differenti modi in cui gli individui e i gruppi lo vivono. Le immagini qualitative, come quelle prodotte, ad esempio, a partire da processi di *Public Participation GIS* [Rambaldi *et al.*, 2009], invece, permettono al ricercatore di rintracciare le informazioni qualitative utili all'analisi. Come si vedrà in seguito, non è solo la presenza di dettagli descrittivi contestuali a caratterizzare tali dati come qualitativi. Il ricercatore sa di trovarsi di fronte a dati contraddistinti da *qualitativeness* quando essi agevolano riflessioni interpretative sul processo che egli intende spiegare. Le storie raccontate e le parole scelte dai rispondenti nella narrazione influenzeranno necessariamente l'interpretazione poiché offriranno non solo informazioni rispetto a un cambiamento fisico e geografico che riguarda il quartiere in cui vivono ma anche possibili nuove interpretazioni di significato sull'evento, sulle loro esperienze, interessi e processo di costruzione della loro identità.

I *Qualitative GIS data*, insomma, aiutano il ricercatore ad esaminare meglio le contraddizioni, le sfumature che portano con sé i dati, attraverso l'identificazione delle informazioni qualitative nei *geodata* e delle informazioni geografiche nel dato qualitativo.

5.5. Database ibrido-relazionale

Sino a qualche anno fa, la scelta dominante assunta dai ricercatori sociali, ai fini di un'integrazione tra CAQDAS e GIS consisteva nel mancato inserimento dei dati qualitativi direttamente nel sistema GIS [Dennis, 2006; McLafferty, 2002; Pavlovskaya, 2002; 2006]. Per capire meglio la natura di tale forma di integrazione, riportiamo un esempio.

Nel corso della sua ricerca sulla condizione economica delle famiglie della Mosca pre e post-sovietica, Pavlovskaya [2002, 281-289; 2004, 329-351] utilizzò i GIS per visualizzare le percentuali sia delle risorse possedute dalle famiglie moscovite, provenienti dalle attività economiche formali e informali, sia del consumo di tali risorse. La ricercatrice utilizzò i dati censuari per analizzare le dinamiche di guadagno e consumo provenienti dalle attività economiche formali. I dati rilevati dalle interviste somministrate alle famiglie servirono, invece, ad analizzare le dinamiche di guadagno e consumo riguardanti le attività economiche informali, quelle cioè non monetizzate, di cui spesso non si trova traccia nelle statistiche ufficiali ma che riguardano molto da vicino la quotidianità delle famiglie (lavori domestici di manutenzione, apporto del lavoro femminile informale all'economia familiare, ecc.).



(Figura 5.1. Rappresentazione di dati riferiti alla condizione economica delle famiglie, 1989-1995, Mosca. Fonte: Pavlovskaya e Martin, 2007, 599)

Entrambi i tipi di dati vennero rappresentati in una mappa costruita tramite GIS, ma i dati raccolti dalle interviste non vennero inseriti direttamente nel sistema. La Pavlovskaya operò, piuttosto, una trasformazione del dato qualitativo

quantificandolo in modo che potesse essere visualizzato congiuntamente ai dati censuari, nel GIS. Questo processo di quantificazione del dato qualitativo presenta un limite [Jung, 2009, 115-132]: nella fase di trasformazione dei dati originari, molte delle informazioni, di cui è portatore il dato qualitativo, e che permettono al ricercatore di interpretare il fenomeno oggetto di studio, vanno persi.

Un'altra tecnica di integrazione tra CAQDAS e GIS che è stata utilizzata dai ricercatori sociali nel corso dell'ultimo decennio è quella multimediale che consiste nel collegare, tramite *hyperlink*, i testi (immagini, audio, video, ecc.) agli oggetti spaziali rappresentati nel GIS. Al-Kodmany [2000], nel corso di una ricerca condotta in collaborazione con il Kirwan Institute for the Study of Race and Ethnicity, Ohio University, collegò foto panoramiche e video, raccolti dagli stessi residenti di alcuni dei quartieri dell'Ohio oggetto di studio, ad un database GIS semplicemente inserendo il *link url* del materiale, presente sul web, nel database GIS [Cieri, 2003; Weiner e Harris, 2003] e conducendo l'analisi dei dati qualitativi rilevati utilizzando i CAQDAS.

Anche quella multimediale è una timida strategia di integrazione che presenta un limite: i dati qualitativi vengono archiviati in un locus esterno al GIS e non appartengono al database principale del sistema. In tal modo, è difficile che i ricercatori possano riutilizzarli per successive analisi.

Come espresso nei capitoli precedenti, le potenzialità analitiche del GIS si manifestano nella possibilità di interrogarlo grazie ai dati archiviati nel database interno al sistema. Perciò, una delle possibili soluzioni, al fine di condurre una CAQ-GIS analysis quanto più possibile rigorosa, sta nell'elaborazione del cosiddetto database ibrido-relazione [*Ibidem*] alla base della quale diventa centrale il processo di codifica.

Occorre, prima di tutto, scegliere il CAQDAS adatto su cui caricare i dati qualitativi. La scelta ricade tra i cosiddetti *Theory Building Software* o software di terza generazione [Mangabeira, 1995, 129-146] che permettono di verificare le

relazioni logiche tra le porzioni di testo riaggregando i segmenti codificati e costituendo una struttura di categorie connesse tra loro⁷¹.

Nel corso del processo di codifica, supportato dal CAQDAS, l'analista attribuisce un'etichetta alle porzioni di testo che egli ha a disposizione, a partire dai dati raccolti. Da un lato, narrazioni, dall'altro, contesti spaziali, espressioni entrambi di significati comuni opportunamente codificati. Il ri-assemblamento di dati, in apparenza, frammentati segue il filo rosso dell'interpretazione. Narrazioni di spazi vissuti, costruzioni sociali situate spazialmente che vengono rilevate attraverso la somministrazione di interviste, nel corso di conversazioni o colloqui in profondità [Holstein e Gubrium, 1997, 113-129]. Tali etichette potranno, poi, essere inserite nel geodatabase GIS, insieme a quelle che identificano il dato quantitativo e spaziale. Seguendo un processo ciclico e iterativo, l'approccio socio-costruzionista si coniuga con quello rigoroso e standard, tipico della cartografia analitica. Da un lato, i CAQDAS privilegiano l'esplorazione induttiva del testo, dall'altro, i GIS permettono di anticipare l'identificazione degli elementi geografici da rinvenire nel testo. Ciò permette di superare i limiti dell'approccio strettamente narrativo⁷², includendo tasselli fondamentali che permettono un'interpretazione più completa.

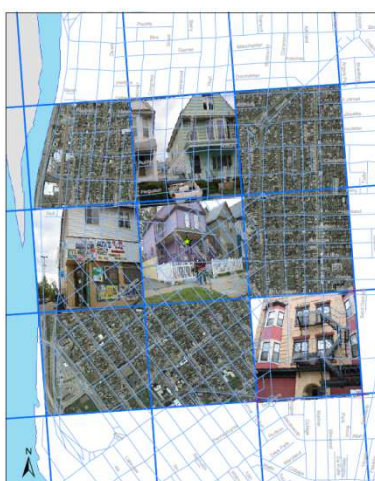
Per dimostrare le potenzialità dello strumento CAQ-GIS, riportiamo l'esempio tratto dalla ricerca di Jin-Kyu Jung [*Ibidem*, 115-132], condotta nei quartieri di Buffalo, New York.

Il ricercatore ha, prima di tutto, costruito quella che definisce la *imaged grid*, la griglia *raster*, costituita da un *layer* nel quale sarà possibile caricare il dato qualitativo⁷³.

⁷¹ *Atlas.ti*, ad esempio, consente di esplorare con agilità i possibili legami tra le categorie (co-occorrenza, sequenza, inclusione, esclusione) e di verificarli attraverso gli operatori booleani, semantici o di prossimità.

⁷² Secondo alcuni studiosi, l'assenza di criteri di classificazione e delle regole riguardanti la selezione del materiale da analizzare e la costruzione delle categorie di analisi, caratterizza il percorso della ricerca qualitativa come poco rigoroso [Losito, 1993]. Oggi, è ormai assodato che anche la ricerca qualitativa si basi su criteri di coerenza e validità [Silverman, 2000, trad. it. 2002].

⁷³ Si rimanda ai capitoli precedenti per l'approfondimento delle tecniche di costruzione di un geodatabase.

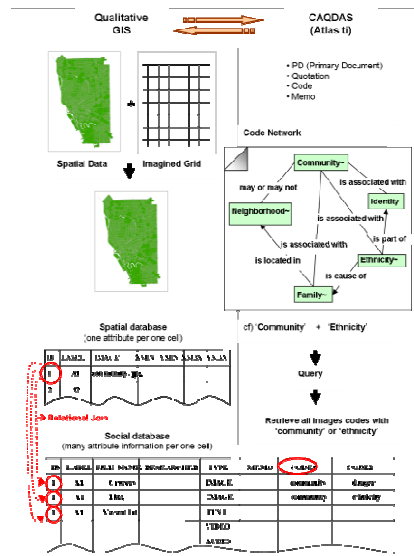


(Figura 5.2. *Imaged Grid*)

Ogni cella della griglia è stata georeferenziata seguendo il procedimento descritto nei capitoli precedenti. Si ricorda che è necessario attribuire ai vertici di ogni cella le coordinate che permettono di localizzare geograficamente l'area geografica interessata. Avvenuta la georeferenziazione, le informazioni riguardanti la localizzazione sono state associate a foto e immagini dei quartieri di Buffalo, raccolte nel corso di uscite sul campo. In questo modo, il dato qualitativo non rimane escluso dal database GIS ma ne fa parte poiché georeferenziato. La griglia, insomma, rappresenta un modo nuovo di raccogliere, archiviare, gestire e visualizzare dati qualitativi che possono essere riutilizzati nel corso di analisi successive.

Tuttavia, la griglia, da sola, non esprime tutte le potenzialità del GIS perché permette di archiviare solo un artefatto qualitativo per cella. Tale limite può essere superato attraverso l'elaborazione del database ibrido-relazionale, comunemente usato nella GIS analysis per collegare le tabelle vettoriali a quelle *raster*.

La struttura del database relazionale permette di costruire relazioni del tipo *uno-a-molti*. Ad un singolo record (riga) della tabella *raster*, può essere associato un record multiplo della tabella vettoriale.



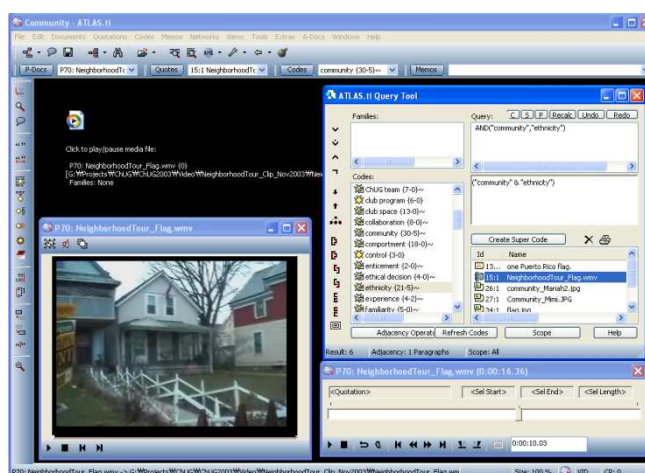
(Figura 5.3. Schema del database ibrido-relazione. Fonte: Jung, 2009, 128)

Si definisce database ibrido-relazionale in quanto comprende sia l'immagine georeferenziata e la rispettiva tabella *raster*, che la tabella vettoriale contenente i codici interpretativi. La tabella dei dati qualitativi si integra, così, a quella dei dati spaziali per mezzo di un unico identificatore: il codice scelto dal ricercatore. I dati possono essere esplorati grazie al pulsante *Identify* il quale, come in precedenza espresso, restituisce:

- i risultati ottenuti dalla tabella *raster*, la quale mostra le informazioni di georeferenziazione e localizzazione dell'immagine qualitativa archiviata nella cella della griglia *raster*, comprese quelle contenenti latitudine e longitudine delle celle (XMIN, YMIN, XMAX, YMAX);
- i risultati ottenuti dalla tabella vettoriale, la quale mostra informazioni riguardanti il tipo di dato, la descrizione dello stesso e le etichette attribuite dal ricercatore.

È chiaro che i codici rappresentano il ponte tra i due sistemi, legame critico tra i CAQDAS e i GIS. Occorre, pertanto, che il sistema di codici, sviluppato nel corso dell'analisi qualitativa supportata dal CAQDAS, coincida con quello riportato nel database ibrido-relazionale. Nel network di codici, elaborato da Jung e riportato in Figura 5.3., quelli utilizzati su *Atlas.ti* (vicinato, famiglia, etnicità,

comunità, identità) coincidono con quelli inseriti nel database ibrido-relazionale. Nel GIS, una stessa immagine potrà essere collegata a due diversi codici (ad es. *house* e *community*). Ciò che importa è che la ricerca dei corrispondenti codici (*house* e *community*) su *Atlas.ti* restituirà tutti i dati che sono stati ad essi collegati e che sono, a loro volta, collegati a quelli presenti nel database ibrido-relazionale GIS. Il ricercatore potrà decidere di iniziare il processo di codifica avvalendosi dell'uno o dell'altro software, in base alle sue preferenze. Quel che è certo è che il processo di popolamento dei campi è iterativo e ciclico, così come prevede il *Qualitative GIS Approach*.



(Figura 5.4. Codifica su Atlas.ti)

La connessione tra CAQDAS e GIS, insomma, avviene lavorando simultaneamente con entrambi i programmi. Il vantaggio sta nel fatto di poter caricare sul geodatabase dati qualitativi georiferibili e riutilizzabili nel corso di successive analisi.

L'approccio di Jung si dimostra molto più convincente di quelli proposti da Pavlovskaya e Al-Kodmany, in quanto rappresenta una soluzione che non prevede il processo di quantificazione dei dati qualitativi e, quindi, la perdita di ricchezza e complessità degli stessi.

5.6. GIS analysis supportata da Atlas.ti

In questo paragrafo, spiegheremo, riportando alcuni esempi di utilizzo, le principali funzioni di integrazione tra *Qualitative* e *GIS data* su *Atlas.ti*,⁷⁴ un software di supporto all'analisi del contenuto di tipo interpretativo che è adatto ad analizzare dati rilevati nel corso delle ricerche di cui riporteremo gli esempi. Ricerche etnografiche in cui il ricercatore osserva, interagisce, annota, elabora concetti e significati e teorizza.

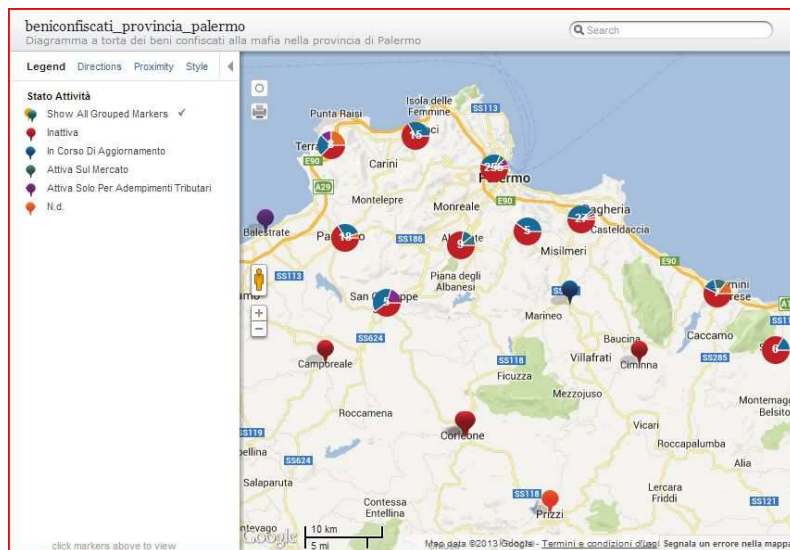
Per mostrare come i dati qualitativi possano essere collegati ai *GIS data*, ci avvarremo anche di *Google Earth*⁷⁵, un software che genera immagini virtuali della terra utilizzando dati satellitari, topografici e foto aeree, memorizzati su una piattaforma GIS.

A differenza di altri CAQDAS (ad esempio, *NVIVO* o *MAXqda*), in cui è possibile inserire esclusivamente il link *url* alla schermata di *Google Earth*, *Atlas.ti* archivia direttamente le localizzazioni all'interno della sua Unità Ermeneutica.

Si farà uso di esempi tratti da uno studio, ancora in corso, condotto dal Dipartimento Scienze Giuridiche, della Società e dello Sport dell'Università degli Studi di Palermo. Nell'ambito del progetto *Imprese criminali: reti di complicità e 'sopravvivenza' delle aziende confiscate alla mafia in Sicilia*, è previsto l'utilizzo dello strumento CAQ-GIS al fine di analizzare sia dati secondari di tipo statistico, cartografico e testuale, sia dati rilevati sul campo, nei territori in cui sono ubicati i beni confiscati alla mafia. L'unità di lavoro, composta da sociologi, metodologi, teorici dell'organizzazione ed economisti, ha previsto di indagare, in una prospettiva multidisciplinare, le complesse reti di relazioni interne ed esterne alle aziende confiscate a Cosa Nostra così come si configurano prima e dopo l'intervento giudiziario.

⁷⁴ Il software è disponibile all'indirizzo www.atlasti.de, da cui è possibile scaricare una versione gratuita.

⁷⁵ Il software in versione desktop è distribuito gratuitamente dalla società Google e scaricabile dal sito <http://www.google.it/intl/it/earth/>.



5.6.1. Archiviare i waypoint in Google Earth

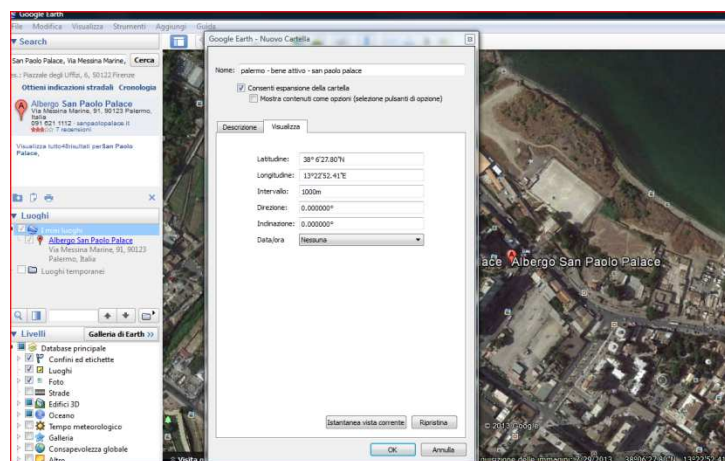
I *waypoint* potranno essere archiviati su *Google Earth* in due modi: caricando sul programma le coordinate che derivano dalla rilevazione GPS o elaborandoli in tempo reale.

Riguardo la prima tecnica, le coordinate di ogni singolo *waypoint* derivanti da una precedente rilevazione GPS, effettuata nel corso di una ricerca sul campo, dovranno essere copiate e incollate nel riquadro ricerca che si trova nella barra laterale del programma.



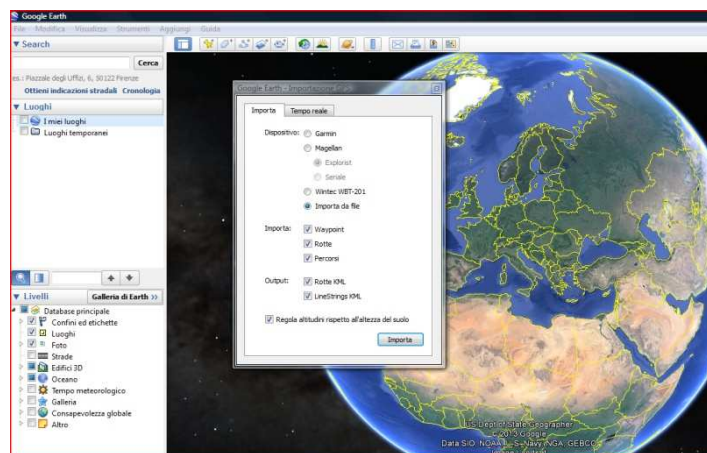
(Figura 5.5. Coordinate nella sezione *Search* su *Google Earth*)

Occorre effettuare l'operazione con precisione, evitando di digitare manualmente latitudine e longitudine, inserendo una virgola e uno spazio bianco che le separi e utilizzando virgolette singole (') per i minuti e virgolette doppie per i secondi ("). Per indicare la direzione, è necessario inserire gli identificatori dei punti cardinali N S E o W (Nord, Sud, Est e Ovest) che possono precedere o seguire il numero della coordinata. In alternativa, è possibile utilizzare il segno meno (-) per indicare l'ovest o il sud e il segno più (+) per indicare una direzione a nord o a est. La prima coordinata rappresenta sempre la latitudine, la seconda, la longitudine. Esempio: 36 55' 12" N, 14 43' 12" E. Cliccando su *Search*, l'immagine verrà centrata, nel piano di visualizzazione 3D, su uno specifico punto geografico. Potrebbe essere utile archiviare i *waypoint* nella sezione *Luoghi* e creare un sistema di cartelle per organizzarli in diversi gruppi, in base alle esigenze di ricerca. Per creare una nuova cartella nella sezione *Luoghi*, occorre cliccare col tasto destro del mouse su *I miei luoghi* e selezionare dal menu a tendina l'opzione *Aggiungi/Cartella*. Per archiviare la localizzazione nella sezione *Luoghi*, occorre cliccare col tasto destro del mouse sulla coordinata presente nella sezione *Search* e scegliere l'opzione *Salva in miei luoghi*. Sarà possibile inserire il *waypoint* archiviato in *I miei luoghi*, nella cartella che preferiamo, trascinandolo col mouse.



(Figura 5.6. Creazione di una nuova cartella)

È possibile, inoltre, importare un intero file KML⁷⁶, contenente *waypoint*, cliccando su *Strumenti*, presente nella barra degli strumenti, e scegliendo *GPS/Importa File* dalle opzioni disponibili nel menu a tendina.

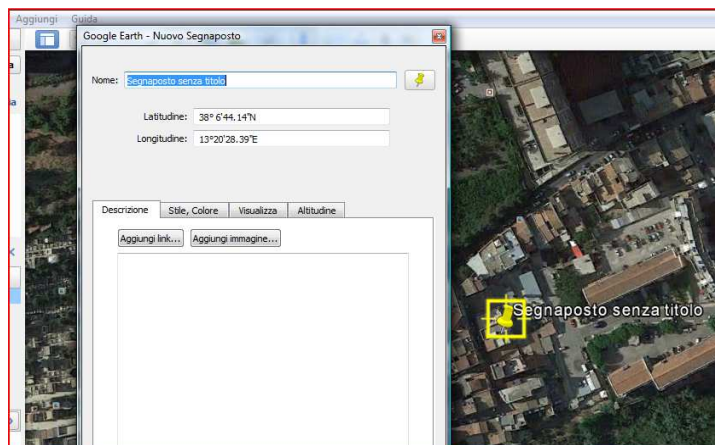


(Figura 5.7. Funzione di importazione delle coordinate su *Google Earth*)

Come già scritto, i *waypoint* possono essere elaborati anche manualmente, da *Google Earth*. Tale tecnica può essere applicata se non si ha a disposizione alcuna rilevazione GPS. Attraverso i pulsanti di controllo di navigazione, il ricercatore potrà spostare l'immagine fino a identificare il punto che desidera contrassegnare. Occorre trovare il giusto equilibrio tra accuratezza della posizione e qualità di risoluzione dell'immagine, effettuando uno *zoom*, sino a che il punto non sia sufficientemente visibile. Centrato il punto da registrare, dalla barra degli strumenti del programma, è possibile cliccare sull'icona a forma di spillo, di colore giallo denominata *Aggiungi segnaposto*. Si aprirà una finestra di dialogo in cui sarà possibile digitare l'etichetta del segnaposto nella stringa *Nome*,

⁷⁶ Il file KML specifica un set di elementi (segnalibri geografici, immagini, poligoni, modelli 3D, descrizioni ed etichette testuali) da visualizzare in *Google Earth*, *Maps* e *Mobile*. Ogni locazione riporta latitudine e longitudine e una serie di altri dati (inclinazione, inquadratura, quota del punto di vista) che rendono la visualizzazione più specifica. Spesso, i file KML sono distribuiti come KMZ, file zip compressi con estensione *.kmz*. Quando un KMZ viene decompresso, al suo interno è presente un singolo file *doc.kml*, contenente tutti gli overlays e le icone referenziate nel KML.

aggiungere *link* e immagini nella sezione *Descrizione*, cambiare stile e colore del segnaposto, dalla sezione *Stile, Colore* e conoscere le coordinate del punto.



(Figura 5.8. Creazione di un segnaposto su *Google Earth*)

5.6.2. Usare *Google Earth* su *Atlas.ti*

Apriamo il software *Atlas.ti*. La prima operazione da effettuare è la creazione della Unità Ermeneutica o l'apertura di una esistente, comprendente al suo interno i dati qualitativi (interviste somministrate sul campo, foto, video, ecc.).

Clicchiamo sull'icona di assegnazione dei *P-Doc*⁷⁷ all'Unità Ermeneutica e scegliamo, dal menu a tendina, l'opzione *New Google Earth PD*⁷⁸.

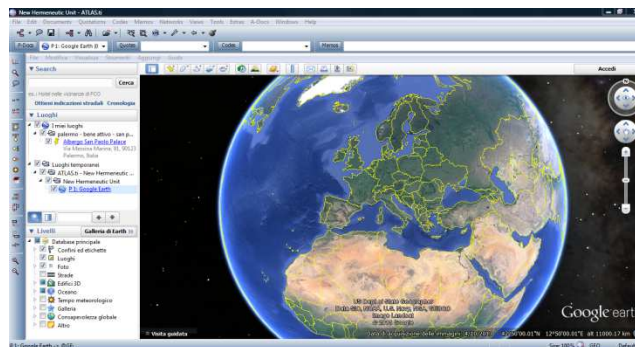


⁷⁷ I *P-Doc* (*Primary Documents*) sono i file (video, audio, testo) che l'utente vuole sottoporre ad analisi. Qualunque file, una volta inserito nell'Unità Ermeneutica, diventa un *P-Doc*.

⁷⁸ I file su cui effettuare l'analisi non vengono fisicamente spostati nell'Unità Ermeneutica. Vengono stabiliti dei collegamenti fra quest'ultima e la destinazione in cui si trovano i dati *Google Earth*.

(Figura 5.9. Apertura di Google Earth in Atlas.ti)

L'opzione scelta creerà un nuovo documento denominato *Google Earth*, che verrà aggiunto alla lista dei documenti del *Document Manager*⁷⁹ e che sarà possibile aprire, come qualsiasi altro documento, all'interno dell'unità ermeneutica. Tutte le opzioni di menu presenti nel software *Google Earth* saranno disponibili su *Atlas.ti*. Le localizzazioni archiviate in precedenza saranno automaticamente importate.

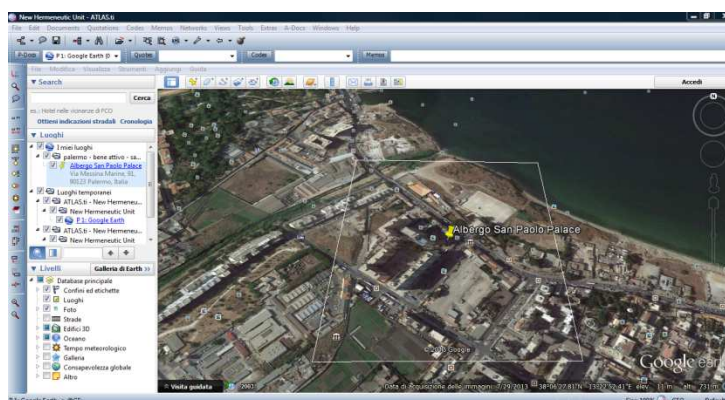


(Figura 5.10. Apertura del documento Google Earth all'interno dell'Unità Ermeneutica di Atlas.ti)

Dopo aver inserito il documento *Google Earth*, il ricercatore potrà iniziare il processo di codifica. Si ritiene che il *waypoint* archiviato sia significativo per il ricercatore che ha deciso di registrarlo. Il ricercatore potrà registrare su *Atlas.ti* il *waypoint* archiviato su *Google Earth* come *Quotation*⁸⁰. Se il *waypoint* di interesse è stato archiviato nella sezione *Luoghi*, si potrà decidere di richiamarlo. Altrimenti, si potrà scegliere di cercarlo inserendo le coordinate o manualmente, come già spiegato.

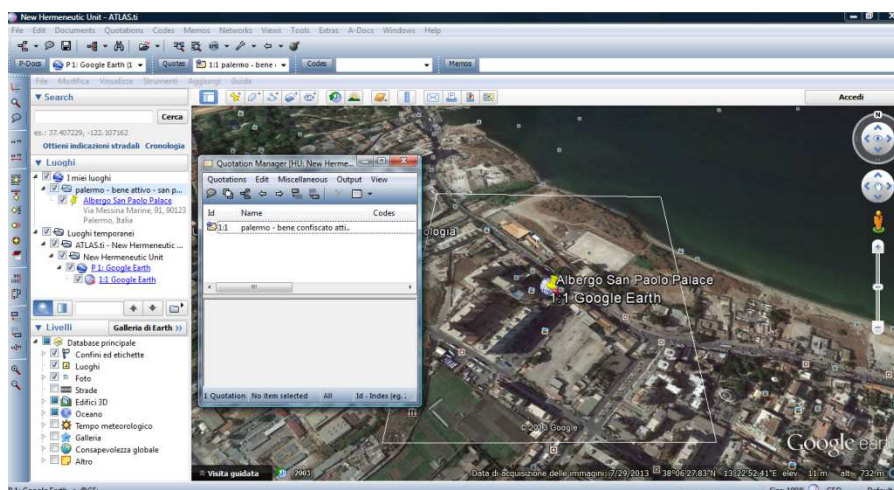
⁷⁹ Il *Primary Document Manager* comprende tutti i documenti che l'utente ha scelto di collegare all'Unità Ermeneutica. Sono visibili nell'area di lavoro cliccando sul menu a tendina denominato *P-Docs* e scegliendo il singolo documento. Il numero (formato P 1: 10.07.03.txt, ad esempio) indica l'ordine, il nome e il formato del documento.

⁸⁰ Le *Quotation* sono estratti di testo che il ricercatore valuta essere significativi tanto da assegnare ad esse almeno un codice.



(Figura 5.11. Richiamo del waypoint Google Earth all'interno dell'Unità Ermeneutica di Atlas.ti)

Dopo aver caricato il *waypoint* di nostro interesse, senza alterare la rappresentazione su *Google Earth*, apriamo il *Quotation Manager*⁸¹ e posizioniamolo accanto alla barra laterale del programma. Clicchiamo su *Create a free quotation*, presente sulla barra degli strumenti del *Quotation Manager*, in modo che una nuova *Quotation*, caratterizzata da un numero identificativo (del tipo, *1:1 Google Earth*) venga aggiunta al *Quotation Manager*. La *Quotation* potrà essere rinominata cliccandovi col tasto destro del mouse.

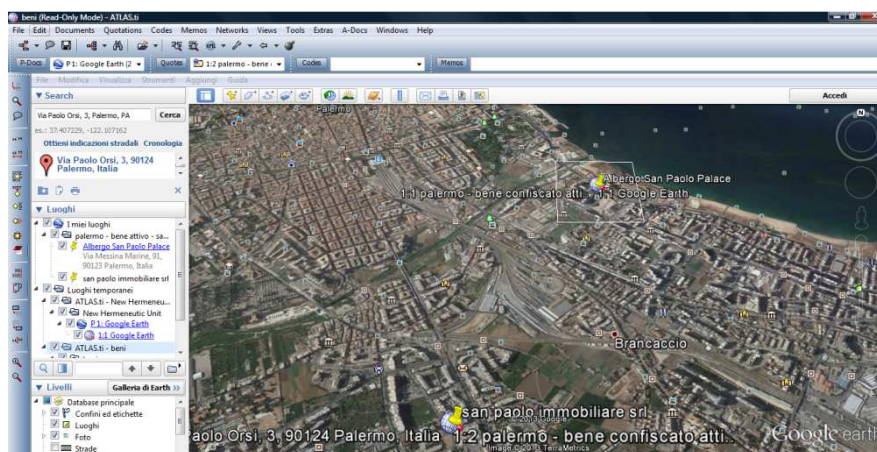


(Figura 5.12. Apertura del documento Quotation Manager all'interno dell'Unità Ermeneutica di Atlas.ti)

Una volta creato e rinominato il *waypoint* e la sua rispettiva *Quotation*, è possibile ripetere la stessa operazione per un altro *waypoint*, finché non abbiamo

⁸¹ Il *Quotation Manager* è l'area in cui vengono salvati e resi visibili i testi che l'utente ha selezionato come significativi o a cui ha assegnato uno o più codici.

esaurito il set. Tutte le *Quotation* create in una sessione verranno incluse nello stesso Documento dell'Unità Ermeneutica⁸². La figura mostra uno *screenshot* a seguito dell'apertura di un documento quotato. I due *waypoint* vengono mostrati sulla mappa con icone a forma di globo.

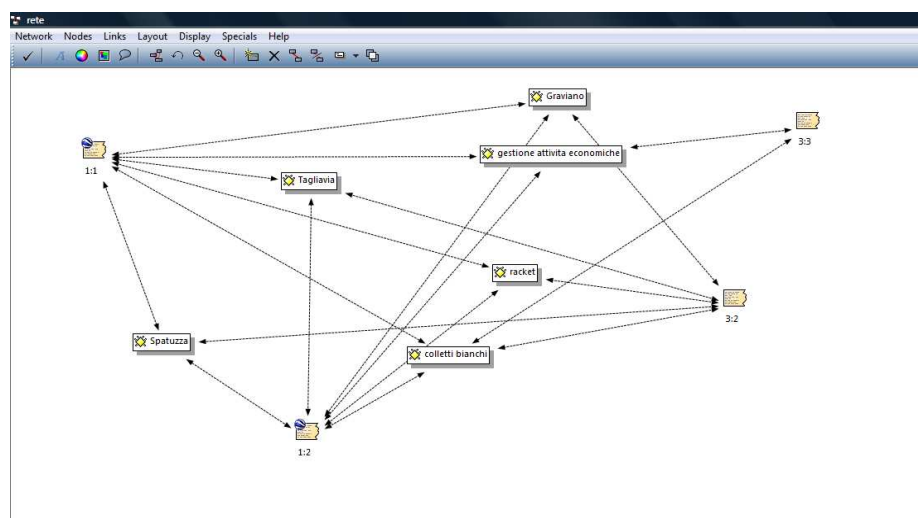


(Figura 5.13. Apertura del documento quotato all'interno dell'Unità Ermeneutica di Atlas.ti)

Considerato che i *waypoints* sono archiviati su *Atlas.ti* come *Quotation*, possono essere usati come qualsiasi altra *Quotation*; nel nostro caso, uno o più codici tematici, associati ad altre porzioni di testo, possono, a loro volta, riferirsi alle localizzazioni. Dallo strumento *Codes*, basta creare un *free code* o sceglierne uno già esistente (*Code by List*) da collegare ad una specifica *Quotation*.

A partire da un elemento selezionato (nel *Quotation Manager*), possiamo costruire una rappresentazione dell'intreccio logico e semantico tra i documenti primari (compreso il *Google Earth Doc*). Clicchiamo con il tasto destro del mouse sulla *Quotation* che ci interessa e scegliamo l'opzione *Open Network View*. Se l'elemento selezionato è stato precedente collegato ad altri, l'output grafico mostrerà tale correlazione.

⁸² Se vogliamo creare un altro documento comprendente un altro set di *waypoint*, è necessario chiudere *Google Earth*, salvare l'Unità Ermeneutica e creare un nuovo documento *Google Earth*. Si è notato che non è possibile creare un nuovo documento senza chiudere *Google Earth* perché ogni nuova *Quotation* verrebbe attribuita al primo documento creato.



(Figura 5.14. Funzione Network view su Atlas.ti)

Non ci sono limiti al numero di volte in cui una citazione può essere collegata a un'altra: una stessa *Quotation* può essere collegata a un documento *Google Earth*, a una foto, a un video, ecc., così come diverse citazioni di testo possono essere collegate a uno stesso documento *Google Earth*.

In questo modo, viene facilitata la navigazione tra i diversi codici e sarà possibile associare ad essi anche le *Quotation* georiferite.

5.7. Participatory Social Research Methods e PPGIS

Le informazioni geografiche e le tecnologie ad esse correlate rappresentano, già da qualche anno, strumenti importantissimi poiché permettono di scomporre gli elementi topografici e, insieme, simbolici, che svelano il senso d'identità condiviso dalle comunità residenti. Sono ampiamente utilizzati, oltre che dai ricercatori sociali, anche da *decision-maker*, durante le fasi più delicate del processo di pianificazione delle risorse e di gestione del territorio.

A coadiuvare il processo di analisi sociale territoriale, il PPGIS [Corbett e Keller, 2005, 91-102; Corbett e Rambaldi, 2009]: un sistema di partecipazione pubblica basato sull'informazione geografica che nasce dalla pratica del *community mapping*. Le numerose esperienze internazionali [De Vera, 2005; Orban-Ferauge, 2011; Rambaldi *et al.*, 2006, 28-35] testimoniano come favorisca, inoltre, la diffusione e il riutilizzo di geodata prodotti dalle stesse comunità. Con il

supporto di *social hacker researcher e mapper*, infatti, le comunità rappresentano se stesse nello spazio, su supporti cartografici o plastici tridimensionali producendo dati che potranno essere inseriti sui portali di dati aperti [Biallo, 2013] concorrendo ad annullare, così, il paradosso di una cultura padrona di milioni di informazioni e che, tuttavia, non sa usarle per il proprio benessere e non sa scambiarsele perché non possiede una mappa mentale abbastanza agile da incorporarle.

Grazie al PPGIS, invece, nomi, simboli, scale e *feature*, legano in modo indissolubile lo spazio geografico all'identità.

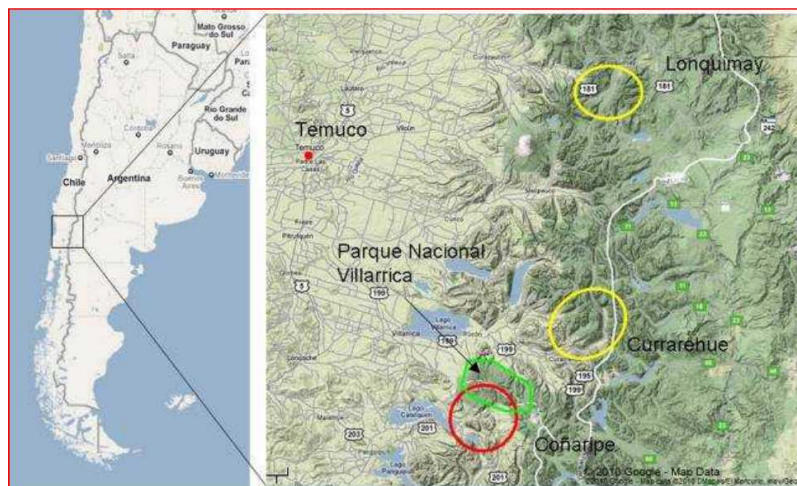
Se si focalizza l'attenzione non sulla *società* come sistema globale, ma sulle relazioni sociali [Donati, 1991], diventa più semplice spiegare come queste ultime siano in grado di generare soggettività sociale in senso solidaristico, attraverso il medium *natura*. Se si osserva che da tali relazioni possa dipendere la possibilità di risoluzione di questioni riguardanti il benessere delle comunità, la pratica del *community mapping* rappresenta una soluzione tutt'altro che tecnocratica.

Il dato geografico riveste il ruolo di infrastruttura abilitante e le tecnologie geomatiche contribuiscono alla realizzazione e messa a punto di servizi innovativi. La geografia è di tutti: è uno dei punti di partenza su cui anche un ricercatore sociale può ragionare per studiare metodi innovativi ed efficaci di raccolta dati.

5.7.1. Mapuche e community mapping

L'esempio riportato è tratto da uno studio realizzato dal CISAI (Centro Interdipartimentale di Studi sull'America Indigena), dell'Università degli Studi di Siena, e dal Centro Studi Americanistici "Circolo Amerindiano" di Perugia, e finanziato dall'Istituto Italo Latino Americano (IILA). Il progetto, conclusosi nel 2011, era volto a ricostruire la storia locale dei *Mapuche* attraverso una ricerca sugli usi del territorio nella coltivazione e raccolta di piante officinali. Il territorio

d'interesse comprende l'area della XIV *Región de los Ríos (Coñaripe)* e della IX *Región de la Araucanía (Currarehue e Lonquimay)*.



(Figura 5.15. Territorio d'interesse dell'indagine CISAI)

Ad animare il progetto *Formazione Istituzionale, Ricerca e Documentazione per lo Sviluppo Agroforestale Sostenibile delle Comunità Mapuche del Cile*, la necessità di coniugare la ricerca pura alla ricerca applicata, coinvolgendo le comunità *Mapuche* nei processi di *decision-making* riguardanti la gestione sostenibile del territorio. Gli analisti coinvolti nel progetto, che realizzano da anni studi nell'area *Mapuche del Cono Sur*, sia dal lato Argentino che dal lato Cileno della Cordillera, aveva individuato, tra gli obiettivi operativi di ricerca, lo sviluppo e la condivisione di un programma sperimentale che consentisse loro di collaborare e scambiare dati eterogenei. A tal fine, il gruppo composito stabilì rapporti di collaborazione con le comunità coinvolgendoli in percorsi di formazione volti a consolidare le loro competenze nell'ambito della documentazione e produzione audiovisiva, del censimento e georeferenziazione delle risorse naturali e socioculturali, da realizzarsi in forma partecipata.

5.7.2. Costruzione partecipata del GIS

Per la costruzione del GIS partecipata, il gruppo si avvale di fonti secondarie di tipo testuale (pubblicazioni, relazioni, ecc.), fonti cartografiche (mappe geografiche, catastali, storiche, foto aeree, immagini satellitari, ecc.), fonti orali

(racconti orali sugli usi tradizionali e ancestrali del territorio)⁸³. Ogni punto segnato sulla mappa venne considerato valido solo se dotato di *forza relativa*; se è supportato, cioè, dai racconti di almeno tre fonti orali indipendenti, da una georeferenziazione avvenuta con partecipazione di tutti i portatori di interesse e avvalorata da videoregistrazioni in loco di racconti e descrizioni.



(Figura 5.16. Community mapping)

La metodologia partecipativa che venne utilizzata nella fase dei censimenti e della georeferenziazione delle risorse è quella tipica del *community mapping*. Il gruppo di ricerca organizzò una serie di riunioni in cui coinvolse le comunità interessate. Durante le prime sedute, i partecipanti elaborarono le prime mappe stilizzate su carta, organizzarono i percorsi da compiere per rilevare i punti rilevanti sul territorio e programmarono le uscite. Nel corso degli incontri, vennero raccolti i primi dati provenienti dai racconti orali, utili a preparare le interviste da effettuare in loco. Le riunioni rappresentarono, di volta in volta, l'occasione privilegiata per confermare la validità della georeferenziazione dei punti sui quali si erano registrate contraddizioni. Ma, considerata la possibilità di confrontare tra dati cartografici, fonti documentarie e fonti orali, il gruppo di ricerca si trovò raramente di fronte ad incongruenze palesi.

Alle riunioni partecipate, i ricercatori scelsero, inoltre, di effettuare anche delle uscite sul campo, in un'unica missione, a metà progetto [Giannelli, 2008, 415-420]. Avvalendosi del supporto di due informatori residenti per ogni comunità, ai quali era stato affidato il compito di scegliere l'area da percorrere, i

⁸³ Il gruppo di ricerca ha tenuto conto anche della dinamica tra gli interlocutori, del contesto in cui i racconti sono stati registrati e dell'intenzionalità dell'interlocutore. Sulla validità e attendibilità delle fonti orali, si rimanda ad altra bibliografia [Andreini e Clemente, 2007].

ricercatori raccolsero foto⁸⁴, video e *waypoints*⁸⁵: un'enorme quantità di dati qualitativi⁸⁶ e quantitativi. Ufficialmente, i ricercatori chiesero agli informatori di accompagnarli nel corso di vere e proprie visite guidate in cui avrebbero potuto apprendere il *mapudungun*. In realtà, lo scopo era quello di ottenere informazioni utili circa i nomi delle specie vegetali più utilizzate e i punti del territorio ritenuti rilevanti per la vita di comunità. Durante la visita, evitarono di indirizzare la conversazione limitandosi, piuttosto, a chiedere chiarimenti sulla corretta posizione del punto indicato dall'informatore. La georeferenziazione venne effettuata per ogni punto in cui era avvenuta un'interazione con l'interlocutore e in cui furono scattate foto.

5.7.3. Sincronizzazione tra *waypoint* e dati qualitativi

I moderni ricevitori GPS sono in grado di tracciare interi percorsi considerati rilevanti dai ricercatori. A conclusione della fase di rilevazione, il gruppo CISAI collegò organicamente ogni foto scattata, ogni risposta ottenuta, ogni fotogramma video ritratto al punto esatto di rilevazione. Una delle tecniche più usate, in questi casi, consiste nel sincronizzare data e orario di tutte le apparecchiature e nell'annotare sul quaderno di campo ogni scelta compiuta durante la rilevazione.

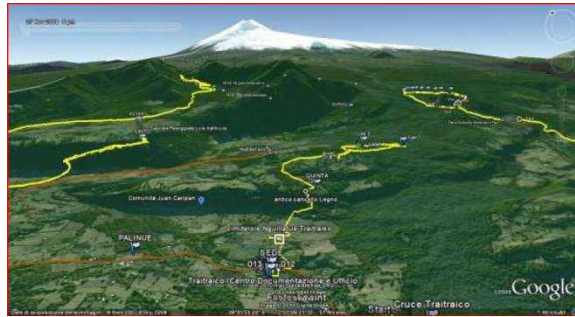
Nel caso di una ricerca di *community mapping*, è auspicabile che i ricercatori facciano uso di software gratuiti e, allo stesso tempo, dotati di funzioni che permettano di condividere i dati via web. La scelta del gruppo di CISAI ricadde su *Google Earth*⁸⁷.

⁸⁴ I ricercatori scattarono tre serie di foto per ogni dettaglio fornito dal referente: dettaglio del referente, immagine con riferimento alle dimensioni e contesto generale in cui il soggetto dell'immagine era localizzato.

⁸⁵ Per la georeferenziazione, venne usato un ricevitore GPS Garmin GPS MAP 60 che permette di attivare la funzione di tracciamento del percorso.

⁸⁶ La durata dell'interazione tra i ricercatori e gli interlocutori andava dalle tre alle cinque ore per ogni percorso.

⁸⁷ In particolare, Google ha rappresentato per i ricercatori una piattaforma di lavoro utile per creare e condividere documenti di interesse per l'intero gruppo.



(Figura 5.17. Percorsi GPS e georeferenziazione *waypoint* tramite Google Earth)

Analizzando i dati qualitativi e qualitativi tramite il CAQDAS *Transana*⁸⁸ e *Google Earth*, i ricercatori hanno concluso che le rappresentazioni del territorio concepite dalle comunità *Mapuche* sono molto diverse da quelle elaborate dai pianificatori. I primi vivono stabilmente nell'area cui si riferiscono i dati rilevati e la utilizzano a fini agricoli, per la loro sussistenza. In Figura 5.18, si riporta un esempio di classificazione ai fini della gestione del territorio elaborato dai pianificatori.

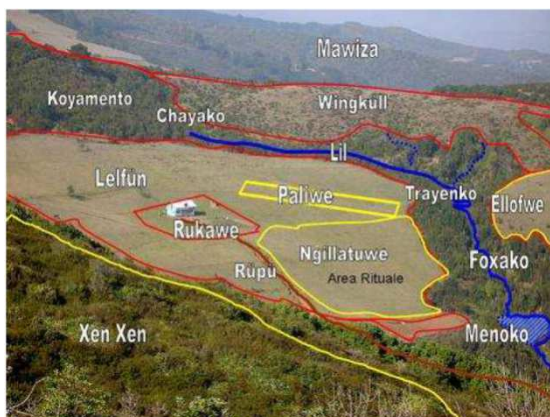


(Figura 5.18. Classificazione del territorio elaborata dai pianificatori)

In Figura 5.19, l'immagine mostra come la famiglia *Mapuche*, che vive nella casa posta al centro dell'area, utilizza una terminologia differente e traccia

⁸⁸ Il software è un semplice CAQDAS che permette ai ricercatori di analizzare dati video o audio in forma digitale. A differenza di *Atlas.ti* non permette l'integrazione dei dati *Google Earth* direttamente nel sistema. Il punto di forza di tale software sta nella possibilità, offerta dalla versione multiutente, di condivisione dei dati in tempo reale. Il database della versione multiutente risiede su di un server, permettendo così ad un gruppo anche numeroso di ricercatori di lavorare contemporaneamente via internet o rete locale ad un unico o più progetti e di comunicare tra loro.

delle suddivisioni differenti che tengono conto delle tecniche tradizionali di gestione del territorio.



(Figura 5.19. Classificazione del territorio elaborata dalla comunità Mapuche)

Ad esempio il *Paliwe* e il *Ngillatuwe* sono aree rituali che non potrebbero essere occupate da coltivazioni. La *Xen Xen* é un'area ricca di piante medicinali e particolari specie commestibili. Attraverso un'attenta indagine etnografica, sono emersi descrizioni e racconti, dettagli utili a ricostruire le pratiche di fruizione dello spazio che garantiscono alla comunità sicurezza e autonomia alimentare, oltre che possibilità di scambio e commercializzazione.

Quella del CISAI è un'esperienza rilevante nell'ambito della *GIS analysis* poiché, anche in questo caso, i GIS hanno costituito uno strumento efficace in grado di mettere a confronto le varie rappresentazioni del territorio e le varie discipline cui i ricercatori coinvolti appartengono: sociologi, antropologi, esperti di diritto, geografi, geologi, linguisti, agronomi. I luoghi vissuti hanno rappresentato, ancora una volta, il terreno di confronto tra punti di vista diversi, contribuendo ad aumentare il livello intermedio di attendibilità della ricerca.

Il *community mapping* non è una pratica che ha finalità esclusivamente conoscitive perché individua anche i metodi di intervento più consoni a promuovere la vitalità economica e sociale di una comunità. Si mantiene lontana dal formulare prescrizioni generali applicabili a tutti i contesti ma tiene conto che, nella vita di comunità, i processi sociali sono complessi per essere standardizzati e troppo singolari per essere considerati in astratto; il loro significato deve essere necessariamente compreso in relazione al contesto specifico di riferimento. Ne

conseguenze che non possono essere identificate regole generali valide per le comunità che vivono in una determinata area, ma sulla base di un'analisi accurata delle dinamiche presenti in ognuna, si possono proporre interventi che rafforzino i fattori di vitalità della comunità stessa. Gli individui che la costituiscono non sono destinatari passivi di interventi calati dall'alto; piuttosto, sono dotati di saperi e risorse basate sulla propria esperienza di vita nella zona comunitaria e, quindi, capaci di avanzare proposte per la soluzione dei problemi urbani⁸⁹.

Lo spazio è uno "scenario" nel quale hanno luogo quotidianamente interazioni comunicative tra soggetti, gruppi; nel quale vengono affinate le capacità comunicative che in parte sono innate, in parte si acquisiscono nella rapidità con cui si consumano i contatti umani. Con il complessificarsi dell'interazione sociale e territoriale si delinea per i soggetti una precisa competenza di destrutturazione dell'esperienza costruita. Il quotidiano, perciò, diviene lo spazio della resistenza rispetto alle pratiche di potere e il luogo della negoziazione tra individuo e società, dell'autonomia e dunque della sovversione di tali pratiche.

Tale esito positivo è frutto di un'efficace unità identitaria favorisce la comunicazione tra i soggetti portatori dello stesso bisogno determinando una situazione di pressione conflittuale tra il gruppo portatore del bisogno e i gruppi che detengono le risorse materiali, culturali e simboliche o gli strumenti per distribuirli come essi ritengono. O il bisogno verrà accolto o respinto. Le definizioni individuali del bisogno tornano a prevalere su quelle collettive e ciascuno utilizzerà come crede le proprie *chance*, per risolvere i suoi problemi [Mela, *ibidem*: 62].

Gli spazi che si fanno luoghi sono la sede prediletta di queste relazioni associative, del loro generare soggettività e di farlo in senso solidaristico.

⁸⁹ È interessante ricordare lo studio di Lévy Strauss sui villaggi Bororo [1955]. La loro struttura rivela un insieme di significati relativi alle pratiche sociali proprie di tale gruppo. Per estirpare le credenze cosmologiche e le pratiche sociali dei Bororo, i salesiani pensarono di costringerli a trasferirsi in villaggi con strutture ortogonali, in modo che perdessero rapidamente il senso delle loro tradizioni.

L'ambiente sociale è il palcoscenico della performance⁹⁰: è in questa che essi cercano alleanze, attingono al patrimonio culturale e simbolico del gruppo a cui appartengono. È nell'ambiente fatto territorio che avviene il confronto/scontro con la classe dominante e le istituzioni egemoniche.

È necessario che le forme fisiche, attraverso le quali il territorio si esprime, siano capaci di cogliere i bisogni, le aspettative dei gruppi e degli individui. Si auspica, perciò, un'urbanità nuova, che tenga conto delle mutate esigenze di controllo e gestione dello spazio da parte dei soggetti e dei gruppi che vogliono sentirsene parte.

Alla luce di quanto scritto, abitare significa modellare il proprio spazio di vita: la riscoperta di un rapporto creativo con l'ambiente. È qualcosa il cui impiego presuppone una saggezza nell'attribuzione di connotazioni simboliche al proprio spazio. Solo così, di fatto, potremo permetterci di imparare l'Altro, sia esso una regione sconosciuta e minacciosa, uno straniero o una proposta diversa. Altrimenti, ogni punto dello spazio diventa equivalente ed estraneo. È evidente il potenziale simbolico di questa concezione: attraverso quest'ordine di considerazioni è possibile legare in modo indissolubile il problema dell'ambiente a quello dell'identità, proponendo una tesi che concili e giustifichi la crisi dell'uno e dell'altra. Il passaggio di questo argomentare è uno:

- Tenere a mente che abitare un luogo non è un'attività fatta di gesti ripetuti ed abituarini anche se questi, per loro intima essenza, aiutano la battaglia contro l'angoscia e l'ignoto. *Perdersi* è l'atto che restituisce al soggetto la misura della strumentalità della sua routine e lo confronta senza filtri all'Altro [Crespi, 1978]. Senza tale confronto, infatti, non esistono più spazi qualificati e gli uomini non riescono a costituirsi come soggetti.

C'è una grande differenza tra una simile soluzione e il procedimento superficialmente culturale di controllo ed appropriazione della realtà tipico delle

⁹⁰ Gli studi sulla *performance*, sviluppatasi a partire dagli anni Settanta e originariamente nati nel contesto degli studi teatrali, indagano sul momento della "messa in scena" delle pratiche culturali, spostando l'analisi sui comportamenti sociali, sul momento dell'enunciazione, sulla dimensione rituale delle pratiche sociali, sulla ricezione e, in particolare, sugli statuti della corporeità [Judith Butler, 1956].

popolazioni evolute; uno scarto qualitativo che spiega il paradosso di una cultura padrona di milioni di informazioni e che tuttavia non sa usarle per il proprio benessere e non sa scambiarsele, perché non ha una mappa mentale abbastanza agile da incorporarle dando loro un senso. L'impossibilità teorica e pratica di un processo conoscitivo così articolato retroagisce negativamente sui gruppi che popolano il territorio. Gli interessi del governo o dei decisori sono, a questo punto, assolutamente antitetici a quelli dei gruppi: laddove questi si formano nell'apparente disordine del tessuto urbano, organizzato secondo percezioni incomprensibili al metro razionale, i primi mirano all'imposizione di canoni e regole tassative. L'immagine razionale si sovrappone alla realtà vitale e la soffoca; ne uccide la creatività, la capacità di restare supporto fertile dell'esistenza della collettività: il sistema di coordinate impersonali diventa il sistema «ufficiale», unico, delle operazioni di trasformazione dello spazio fisico. La costruzione dell'ambiente diventa un'operazione comandabile.

La pratica del *community mapping* si fonda sul camminare e perdersi nello spazio in cui ci è dato di vivere: lo spazio diventa un racconto lontano dal testo pianificato e leggibile [De Certeau], dall'ordine costituito. Se questo non avviene, la vita urbana potrebbe lasciar riaffiorare ciò che i progetti dei pianificatori hanno escluso. Più la città-concetto è organizzata, più si deteriora; più esclude l'ignoto predisponendo lo spazio, più ci si allontana dall'inconoscibile che favorisce, invece, la scoperta e la maturità. La predisposizione degli spazi altrui senza consultazione si inserisce, quindi, nella più ampia riduzione del mondo a *res extensa*. Tutto è riportabile sulla croce sacrificata degli assi cartesiani. Nel sistema di orientamento *del contesto*, dell'uno rispetto all'altro, invece, si può comprendere ciò che avviene in un luogo a partire dal proprio.

Se la mancata sistemazione dello spazio secondo le dinamiche inclusive genera paura, allora occorre proprio aggiornare le strategie in chiave creativa perché lo spazio diventi un luogo dove vivere e *incontrarsi bene*.

CAPITOLO SESTO SVILUPPI FUTURI DELLA GIS ANALYSIS

6.1. Geographic Information Systems e Social Network Analysis: Spatial is Special

Una delle funzioni principali della GIS *analysis* è quella di spiegare la natura *non uniforme* delle distribuzioni dei fenomeni nello spazio. La necessità di scoprire le regole che definiscono l'organizzazione di tali disposizioni apparentemente casuali, oggi più di prima, spinge l'analista GIS a far uso di strumenti che permettano di analizzare meglio le morfologie tipiche delle aggregazioni spaziali dei fenomeni sociali [Anderson *et al.*, 1999].

La Social Network Analysis studia le relazioni tra più unità interdipendenti, dette *nodi*, mettendo a fuoco i modelli di rete [Trobis e Milia, 2011]. La sua integrazione con la GIS *analysis* si colloca in uno snodo teorico e metodologico importante passando attraverso alcuni dei più importanti dilemmi teorico-interpretativi della sociologia: riduzionismo/olismo, micro/macro, qualità/quantità. Il *fertile connubio* [Trobis, 2005] GIS-SNA si situa a un livello meso e permette di rispondere alle seguenti questioni: «Quali strutture caratterizzano l'organizzazione delle relazioni sociali?», «Dove si situano e come vengono prodotte tali strutture?», «Che forma hanno?», «Quali sono i confini geografici delle reti?».

L'identificazione dei confini delle reti può risultare problematica. «Tali confini sono confini di senso, simbolici e culturali, definiti da norme e valori che non possono essere semplicemente derivati dalla struttura delle relazioni sociali, né, come altri aspetti della macrostruttura, possono essere tutti ricostruiti e spiegati, nella loro autonomia dinamica, tramite l'aggregazione di relazioni elementari» [Mutti, 1996, 15]. La forma delle relazioni viene, in genere, esaminata da due prospettive complementari: una, interessata a misurarne la forza, l'altra occupata a rintracciare e classificare le posizioni, individuali o collettive, all'interno della rete [Trobis, 2001]. Ogni posizione occupata dal nodo individua

le caratteristiche formali di un reticolo sociale che può essere analizzato seguendo tre criteri [*ibidem*]:

- Grafico-formale: si basa sulla forma complessiva del reticolo;
- Posizionale o topologico: si basa sulla posizione dei singoli nodi;
- Spaziale: si basa sul calcolo delle distanze fra i nodi.

La GIS *analysis* e la SNA potrebbero essere considerate due aree di riflessione metodologica diverse eppure, come riconoscono gli analisti di entrambi gli approcci, spesso i *network* si formano entro i limiti di specifiche strutture spaziali. Tale prospettiva è avvalorata da una lunga tradizione di ricerca empirica [Bossard, 1932; Stewart, 1941; Zipf, 1949; Morrill e Pitts, 1963; Irwin e Hughes, 1992].

I fatti sociali sono *situati* [Abbott, 1997, 1152]. E proprio negli ultimi anni, il contesto spaziale nel quale si manifestano ha riguadagnato l'attenzione degli scienziati sociali che impiegano la SNA per lo studio dei fenomeni sociali. Il concetto di *Collective Efficacy* [Morenoff *et al.*, 2001], secondo il quale i meccanismi di fiducia tra i soggetti sono generati anche dalla loro vicinanza fisica, ha spinto i ricercatori ad analizzare le reti che nascono e si strutturano in uno specifico contesto geo-spaziale. Sampson, a tal proposito, ridefinisce il concetto di vicinato come «Nodes in a larger network of spatial relations» [*Ibidem*, 158].

L'utilizzo, da parte dei ricercatori SNA, di strumenti di analisi che spiegano la natura di tale condizionamento reciproco (*Nodes-Spatial*) [*Ibidem*], li spinge ad incorporare gli *Spatial Effect* nei loro modelli e a considerare i *neighborhoods as nodes*. *Reciprocità* [Holland e Leinhardt, 1981], *predisposizione alla vicinanza transitiva* [Hollande e Leinhardt, 1971; Frank, 2005; McPherson *et al.*, 2001], sono concetti che spiegano la natura di tale dipendenza e che nascono dall'interazione tra percorsi di ricerca diversi.

Il testo di Wasserman e Faust⁹¹ [1994], nel quale si ritrova una sintesi dei concetti riguardanti quella che oggi potrebbe essere definita la classica SNA,

⁹¹ Occorre ricordare anche i testi introduttivi di Degenne e Forsé [1999].

comprova la natura multidisciplinare che ha caratterizzato tale approccio. Poco dopo la pubblicazione del testo, i suoi principali concetti analitici sarebbero stati meglio descritti, le tecniche meglio affinate. Brandes e Erlebach [2005], Carrington *et al.* [2005], Doreian *et al.* [2005], Trobia e Milia [2011], a tal proposito, presentano un quadro chiaro sul suo percorso di evoluzione che giunge sino allo sviluppo di tecniche di modellazione statistica per i dati di rete [Jones e Handcock, 2003; Morris, 1991] e di analisi delle cosiddette *strutture spaziali latenti* [Hoff *et al.*, 2002; Handcock *et al.*, 2007; Krivitsky e Handcock, 2008]. Altri hanno sottolineato come, per analizzare le strutture dei *network situati spazialmente*, sia necessario ricorrere a precise tecniche statistiche che permettano la fusione di entrambe le tradizioni analitiche [Butts, 2003; Hipp e Perrin, 2009]. I due approcci presentano una forte sovrapposizione concettuale. Ad esempio, il grado di *omofilia* [McPherson *et al.*, 2001; Aral *et al.*, 2009], che indica la tendenza degli attori a formare legami sociali con chi è più simile a loro, assume nella GIS *analysis* numerose altre etichette. Una di queste è la cosiddetta *autocorrelazione spaziale* [Tobler, 1970], secondo la quale il comportamento di un individuo tende ad essere simile a quello dei suoi vicini e la distanza sociale spesso coincide con quella fisico-spaziale [Newman, 2002]. Tuttavia, l'integrazione formale tra SNA e strategie analitico-spaziali resta relativamente ancora poco affrontata in letteratura [Anselin, 2002].

6.2. Spatially embedded network e autocorrelazione spaziale

Gli individui scelgono i luoghi in cui abitare, in cui passare il proprio tempo libero, in cui lavorare, in cui accedere alle risorse di cui hanno bisogno. Identificare tali localizzazioni può aiutare lo scienziato sociale a dar forma a modelli spaziali utili all'interpretazione delle disposizioni relazionali [Blacksher e Lovasi, 2011].

Come già scritto, una delle questioni più spinose che riguarda lo studio dei *network* è data proprio dalla difficoltà di determinare i confini fisici della popolazione oggetto di studio [Marsden, 2011]. Esiste una varietà di strategie volte ad identificarli ma si ignora spesso quanto possano limitare l'osservazione in

maniera inappropriata. Quando la linea definitoria appare coincidere perfettamente con la realtà sociale, ecco che quest'ultima finisce col discostarsene, ad un'analisi più approfondita.

Secondo numerosi autori [Hipp, 2012; Flowerdew *et al.*, 2008; Lee *et al.*, 2005; Openshaw, 1984], la combinazione tra *network analysis* e approccio analitico spaziale offre una soluzione alternativa anche a tale tipo di questione.

Per definire uno *spatially embedded network*, utilizzano alcuni dei termini presi in prestito dalla teoria matematica dei grafi [Frank, 2005]. Essa studia i *network* come set di *entità* (persone, gruppi o organizzazioni, artefatti, concetti), oggetti caratterizzati da relazioni discrete, e schematizza i processi attraverso cui queste ultime si formano, non in base al loro contenuto ma in base alle loro proprietà formali.

Definito il grafo G come una struttura relazionale consistente di due elementi - un *set* di entità, chiamate *vertici* o *nodi*, e un *set* di *edge* o *legami*, indicanti la relazione tra le entità - esso può esser descritto formalmente nel seguente modo [Butts *et al.*, 2011, 228]:

$$G = (V, E)$$

dove V è il *set* dei *vertici* ed E è il *set* di *edge*.

I grafi si manifestano secondo diverse forme e varietà che dipendono anche dal tipo di relazione che caratterizza i loro *set* di legami. La struttura del grafo cui si farà riferimento qui sarà composta da nodi *spatially embedded* in S . La distanza (d) tra le localizzazioni dei due nodi sarà:

$$d(l(v_i), v_j)$$

dove S è lo spazio in cui sono localizzabili i vertici v_i e v_j ed l è l'operatore di localizzazione [Ibidem, 229].

La *Prima Legge di Tobler* [1970], conosciuta anche come *Prima Legge della Geografia*, stabilisce che ogni entità localizzata nello spazio geografico

ponga in essere una relazione. Ogni entità è legata al resto ma le entità vicine stabiliscono una relazione con maggiore probabilità rispetto a quelle tra loro lontane. Ciò che permette di rilevare la somiglianza tra le entità tra loro vicine è la misura della già citata *autocorrelazione spaziale*. La prossimità spaziale di valori tra loro simili indica un'autocorrelazione spaziale positiva. Essa si può ottenere applicando le diverse tecniche sviluppate nell'ambito della *Spatial Cluster Analysis* [Jacques, 2008] e che si pongono, come obiettivo, l'individuazione di *cluster*, cioè di aree caratterizzate da un'elevata densità del fenomeno in esame.

La *Nearest Neighbour Analysis* (analisi di vicinato) è un esempio di *spatial cluster analysis*. L'indice di vicinato o NNI (*Nearest Neighbour Index*) è il rapporto tra la distanza di vicinato osservata (\bar{D}_O) e la distanza attesa (\bar{D}_E), calcolata in base a un *set* di punti scelti a random dalla procedura. L'indice raggruppa le coppie di punti che superano il test probabilistico [Gelman *et al.*, 2000]. Se il valore dell'indice è maggiore di 1, viene generato il *cluster*.

$$NNI = \frac{\bar{D}_O}{\bar{D}_E}$$

Tra le tecniche di *clustering* applicabili a oggetti poligonali, di fondamentale importanza l'*indice di Moran* [Anselin, 2000] che misura il livello di concentrazione dei fenomeni nello spazio, in base alla distanza e al grado di somiglianza/dissomiglianza tra le aree tra loro confinanti. La *I di Moran* assume valori positivi in presenza di aree tra loro simili e valori negativi in presenza di aree tra loro dissimili. Il suo valore varia da -1 a +1; più si avvicina allo 0 e più il fenomeno è casuale.

La maggior parte dei software GIS permette di integrare ai *tool* presenti di *default*, strumenti aggiuntivi che facilitano il calcolo delle statistiche spaziali più importanti e la visualizzazione dei risultati analitici, sottoforma di tabelle, mappe o grafici. ArcGIS, ad esempio, mette a disposizione degli utenti lo *Spatial Statistics Toolbox* dal quale è possibile calcolare sia l'NNI e che l'*I di Moran* [Mitchell, 2005].

6.3.1. Dengue fever: un esempio di NNI Spatial Analysis

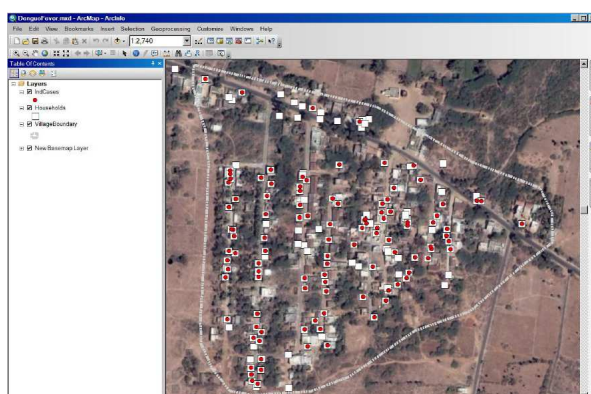
La febbre *dengue* è una malattia molto diffusa nel Sudest Asiatico e in America Centrale. Ogni anno, la contraggono circa 100 milioni di persone.

In questo paragrafo, si mostrerà come usare ArcGIS al fine di comprendere se i casi di febbre *dengue*, registrati in un ipotetico villaggio, si clusterizzano seguendo particolari *pattern*.

Si apra il software ArcMap e si carichi l'ArcMap document *DengueFever.mxd*, accessibile all'indirizzo⁹²:

<https://www.dropbox.com/sh/pr3z6pskrffw20w/rGimafB57c/SpatialAnalysis>.

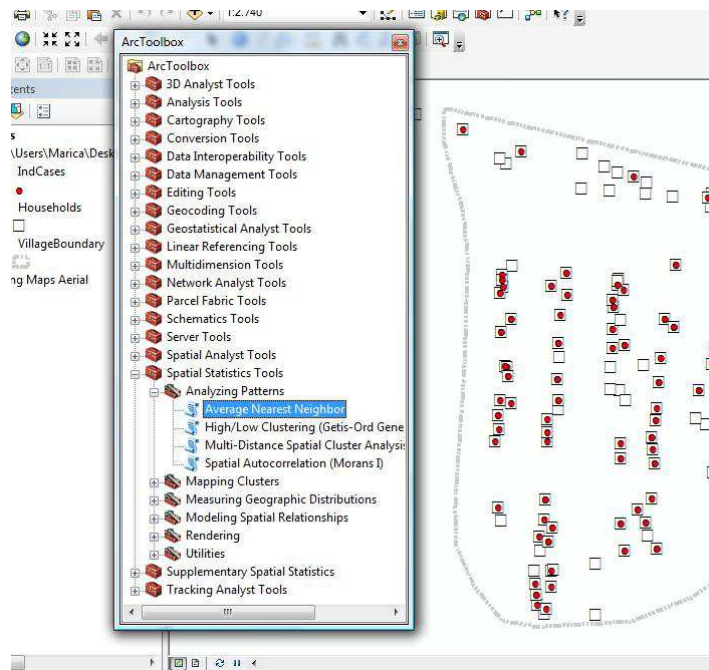
I quadrati bianchi che comprendono un cerchio rosso sono le abitazioni in cui vivono i soggetti che hanno contratto la *dengue*, entro un periodo di 35 giorni.



(Figura 6.1. Abitazioni in cui vivono i soggetti che hanno contratto la *dengue*)

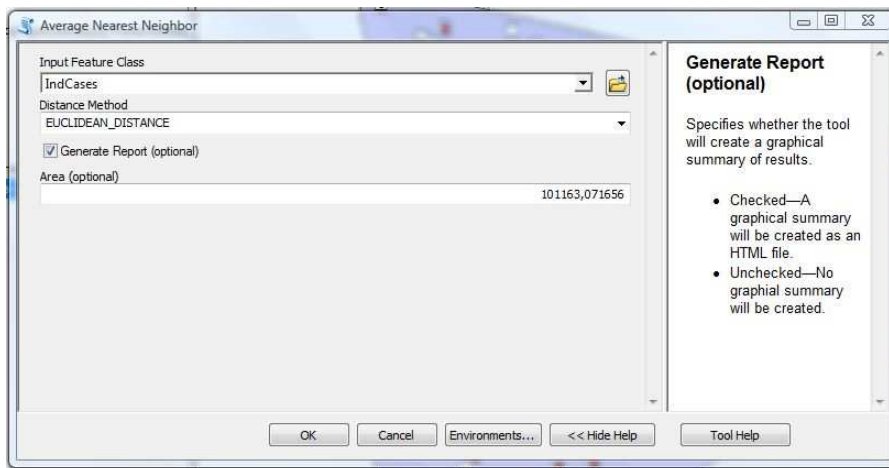
Dall'*ArcToolbox*, si scelga il set di strumenti denominato *Analyzing Pattern* e si apra l'*Average Nearest Tool*.

⁹² Il documento comprende già il geodatabase precedentemente costruito.



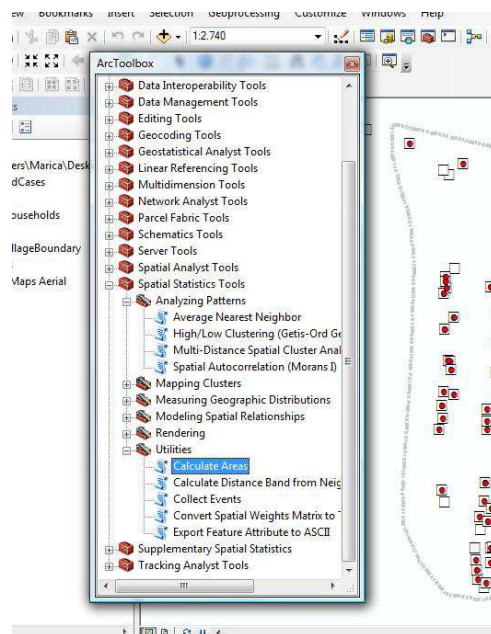
(Figura 6.2. ArcToolbox. Scelta dello strumento *Average Nearest Neighbor* dal set *Analyzing Patterns*)

La finestra di dialogo *Average Nearest Neighbor* richiederà di inserire il valore dell'area d'analisi utile a individuare la \bar{D}_E .



(Figura 6.3. Finestra di dialogo *Average Nearest Neighbor*)

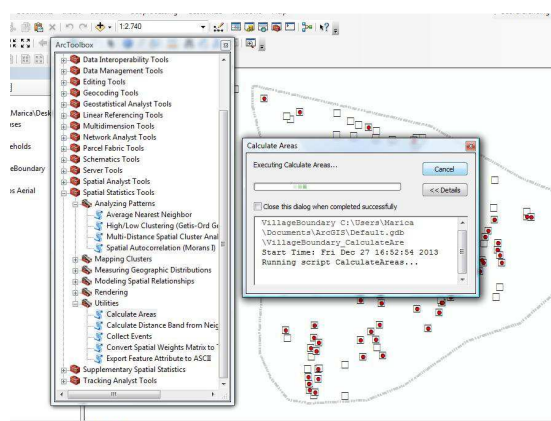
Per il momento, si riduca ad icona la finestra e si apra lo strumento *Calculate Area*, al quale è possibile accedere dalle *Utilities* dello *Spatial Statistics Toolbox* presente nell'*ArcToolbox*.



(Figura 6.4. Apertura strumento *Calculate Areas*)

Tale strumento permette di calcolare l'area di interesse della ricerca: quella che comprende tutte le abitazioni e delinea il poligono dell'intero villaggio. La *feature class* di *input* (*VillageBoundary*) era già stata definita in precedenza nel processo di costruzione del geodatabase⁹³. Basterà richiamarla, come *input*.

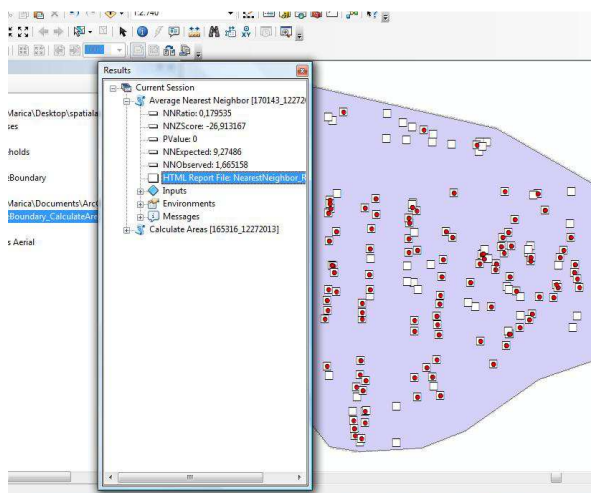
Dopo aver cliccato su OK, lo strumento *Calculate Area* genererà la *feature class* *VillageBoundary*.



(Figura 6.5. Generazione *Feature Class Village Boundary*)

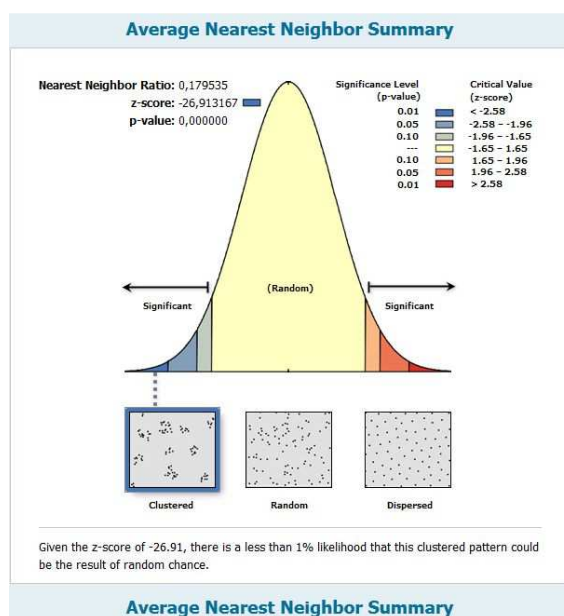
⁹³ Tutti i file che compongono il geodatabase sono accessibile all'indirizzo sopra indicato.

Si apra la relativa tabella degli attributi, si copi il valore dell'area del poligono e lo si incolli nella finestra *Average Nearest Neighbor*.



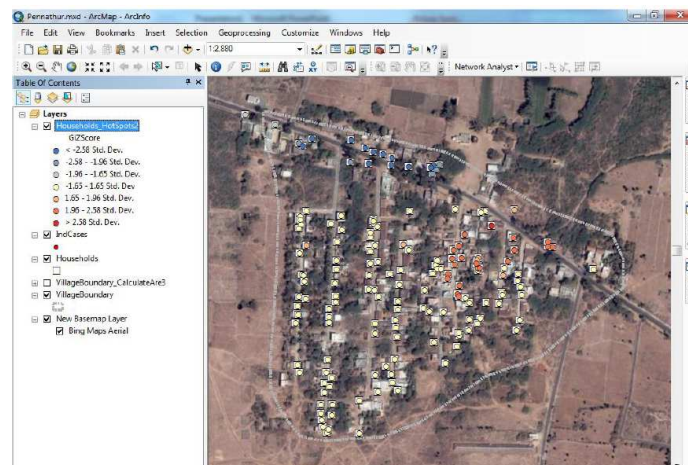
(Figura 6.6. Comando *Results* accessibile dal menu *Geoprocessing*)

Si acceda al comando *Results* attraverso il menu *Geoprocessing*. In questo modo, l'analista potrà visualizzare un file HTML comprendente un *report* e un grafico che mostrerà la clusterizzazione dei casi *dengue*.



(Figura 6.7. *Average Nearest Neighbor Summary*)

Lo z -score (-26) è la *deviazione standard*. Più si avvicina allo zero e più la media delle osservazioni ricadrà al centro della curva normale. Quando lo z -score è negativo e ricadrà, quindi, in corrispondenza della coda sinistra della curva (blu), il *pattern* spaziale sarà più clusterizzato rispetto a un corrispondente *pattern random*. Se lo z -score è positivo, ricadrà in corrispondenza della coda destra della curva normale (rosso) e il *pattern* spaziale sarà più disperso rispetto a un *pattern random*.



(Figura 6.8. Clusterizzazione abitazioni in cui vivono i soggetti che hanno contratto la *dengue*)

I cluster visibili in figura 6.8. si sono sviluppati a seguito di un flusso di comunicazioni e di relazioni che hanno determinato la diffusione della *dengue*. Tenderanno a conservarsi nel tempo soltanto nella misura in cui il flusso continuerà a verificarsi, creando canali relazionali consolidati.

6.4. I terabyte di Tobler

Wikipedia è tra i primi dieci siti al mondo più visitati su *Internet*. È composto da 25 edizioni diverse, una per ogni lingua, e contiene quasi 3 milioni di articoli nella sola edizione inglese. Ogni articolo (o *entità*) dà origine ad una o più relazioni con altri articoli, attraverso i *link*. Tali relazioni sono codificate dai *contributor* (o *wikipedian*) che elaborano gli articoli.

Per verificare la *Prima Legge di Tobler*, Hecht e Moxley [2009] hanno studiato le relazioni unidirezionali tra gli articoli di *Wikipedia*, quelle cioè che generano il

link da un articolo, elaborato in una specifica lingua e in un certo contesto geo-spaziale, all'articolo dedicato allo stesso argomento, elaborato in lingua diversa e in un contesto geo-spaziale diverso, localizzato a una certa distanza dal primo. L'esperimento condotto dai due ricercatori su tale enorme mole di dati è considerato il più empirico tra quelli che hanno tentato di dimostrare la Legge di Tobler [Ortega *et. al*, 2007].

I ricercatori, che definiscono lo studio come *iperlinguistico*, hanno sviluppato un set di API *java*, chiamati WikAPIdia. I WikAPIdia rendono possibile l'accesso spazio-temporale a un certo numero di edizioni in lingua diversa di Wikipedia. Permettono, inoltre, di produrre un certo numero di grafi di analisi (WAG o *Wikipedia Article Graph*) e di ricostruire i processi di formazione del linguaggio naturale attraverso strutture ricorrenti. La fonte di tali dati è costituita da diversi database XML, messi a disposizione dalla Fondazione Wikipedia. Attraverso WikAPIdia, tali file XML restituiscono uno *snapshot*, in tempo reale, dello stato in cui Wikipedia si trova nel momento in cui sono scaricati. Per la presente ricerca, gli autori hanno usato i pacchetti XML riferiti a 22 differenti versioni di Wikipedia e hanno considerato circa 100 mila articoli per ogni *download XML* effettuato⁹⁴. Nonostante la gran quantità di dataset scaricati, solo alcuni sono stati utilizzati nel corso della ricerca. Li riporta la Tabella 6.1.

⁹⁴ Va considerato che gli autori hanno effettuato dei *download* ravvicinati nel tempo per minimizzare gli errori dati da effetti esterni. È un'analisi che va condotta attraverso un PC potente e nel corso di diversi giorni successivi.

Language	Vertices (Articles) = $ V $	No. Edges (Links) = $ E $	Spatial Articles = $ V_{spatial} $
Catalan	141,277	3,478,676	13,474
Chinese	203,824	5,566,490	14,177
Czech	112,057	3,089,517	8,599
Danish	97,825	1,714,025	7,118
Dutch	497,902	9,679,755	103,977
English	2,515,908	76,779,588	174,906
Finnish	208,817	3,782,563	11,559
French	716,557	20,578,831	67,042
German	827,318	21,456,176	85,906
Hungarian	120,850	3,009,814	6,939
Italian	516,120	14,968,632	67,433
Japanese	532,496	20,946,112	21,621
Norwegian	193,298	3,774,509	16,607
Polish	555,563	12,678,608	58,367
Portuguese	437,640	8,577,435	79,844
Romanian	118,345	1,434,939	20,349
Russian	341,197	7,762,322	30,346
Slovakian	102,089	1,931,138	7,708
Spanish	443,563	12,576,477	42,431
Swedish	295,605	5,555,219	18,816
Turkish	120,689	2,260,241	5,431
Ukrainian	131,297	1,743,304	4,692
TOTAL	9,230,237	243,344,371	867,342

(Tabella 6.1. Statistiche descrittive del WAG e numero delle entità spaziali (articoli) per ognuna delle edizioni di Wikipedia incluse nello studio. Fonte: Hecht e Moxley [2009]).

Il WAG estratto da ogni WikAPIdia è un grafo che ha, ai suoi *vertici* (o nodi), gli articoli della edizione di Wikipedia. Gli *edge* (o collegamenti), invece, rappresentano le relazioni tra gli articoli [Clauset *et al.*, 2009].

Secondo la terminologia utilizzata nell'ambito della teoria matematica dei grafi, il WAG è un *multigrafo sparso diretto* perché i suoi *link* sono caratterizzati da una direzione (un *link* va da un articolo a un altro) e ogni nodo è connesso ad un numero relativamente piccolo di altri nodi.

I *link iperlinguistici*, codificati dai *wikipedian*, sono le voci multilinguistiche, presenti nella barra sinistra di ogni pagina. Passando in rassegna tali *link*, WikAPIdia è in grado di riconoscere, ad esempio, se l'articolo in lingua inglese "*Psychology*" e quello in lingua tedesca "*Psychologie*" si riferiscono entrambi allo stesso concetto e di estrapolarne i *geotag*. Questi ultimi riferiscono su latitudine e longitudine⁹⁵ del *wikipedian* che ha prodotto l'articolo.

Se si considerano i tre articoli spaziali (A, B e C) e si assume che la distanza tra la posizione in cui è stato elaborato l'articolo spaziale A e quella in cui è stato elaborato l'articolo spaziale B sia minore della distanza tra la posizione in cui è stato prodotto l'articolo spaziale A e la posizione in cui è stato prodotto

⁹⁵ WikAPIdia assume che i tag si riferiscano al sistema di riferimento World Geodetic System 1984 (WGS1984).

l'articolo spaziale C, ci si può aspettare che l'articolo spaziale A dia luogo a una relazione unidirezionale con B, piuttosto che con C. Se la legge di Tobler è valida, $P_{relation}(A,B)$ si verifica con più probabilità rispetto alla $P_{relation}(A,C)$, dove $P_{relation}$ è la probabilità che l'articolo spaziale a contenga un *link* o una relazione unidirezionale verso l'articolo spaziale b . Considerando che la Prima Legge di Tobler stabilisce che «Ogni elemento è correlato a qualsiasi altro elemento nello spazio geografico» [Tobler, 1970], è possibile ipotizzare inoltre che:

$$P_{relation}(A,C) > 0,$$

anche se A e C sono entità separate da una distanza rilevante.

Per testare tale ipotesi, gli autori hanno esaminato tutte le coppie di articoli spaziali (A, B), escludendo i casi di identità (A, A) e registrando, per ogni coppia, la linea di distanza, in base alla geolocalizzazione dell'articolo.

Dall'analisi dei dati, i ricercatori hanno riscontrato che la classe minima di distanza tra le entità, affinché si stabilisca una relazione, debba essere pari a 50 Km. Tale asserto si traduce formalmente indicando che la probabilità complessiva P_d che si stabilisca un *link* tra due articoli spaziali separati dalla distanza d è data dal numero di relazioni esistenti in una specifica classe di distanza diviso il numero totale di possibili *link*:

$$\frac{\sum_{i=1}^{|PAIR_d|} relation(PAIR_d)}{|PAIR_d|} = P_d$$

dove: $PAIR_d$ è il *set* di tutte le coppie di articoli spaziali in una classe di distanza d ; *relation* verifica l'occorrenza o meno di un *link* che si dirige dal primo articolo spaziale al secondo; e $|PAIR_d|$ è il numero di coppie di articoli spaziali. È importante notare che (A,B) e (B,A) sono considerate coppie diverse poiché A può essere relazionato a B, ma non viceversa. Pertanto, il numero dei possibili *link* deve essere calcolato considerando entrambe le coppie.

La difficoltà nel condurre tale tipo di analisi si manifesta soprattutto nella computazione di un grande numero di entità e nella conseguente potenziale mancata comparabilità tra edizioni differenti, le quali sono caratterizzate da un numero diverso di interazioni ed entità spaziali. Basti pensare che l'edizione inglese potrebbe essere composta da circa 30 bilioni di interazioni. È il motivo per cui i ricercatori hanno deciso di trattare i Wikipedia costituiti da un numero elevato di interazioni dividendo queste ultime in sottoinsiemi random di 50 mila. In questo caso, la misura di *linkabilità*, la probabilità cioè che si verifichi un *link* da un articolo random a un altro, è data da:

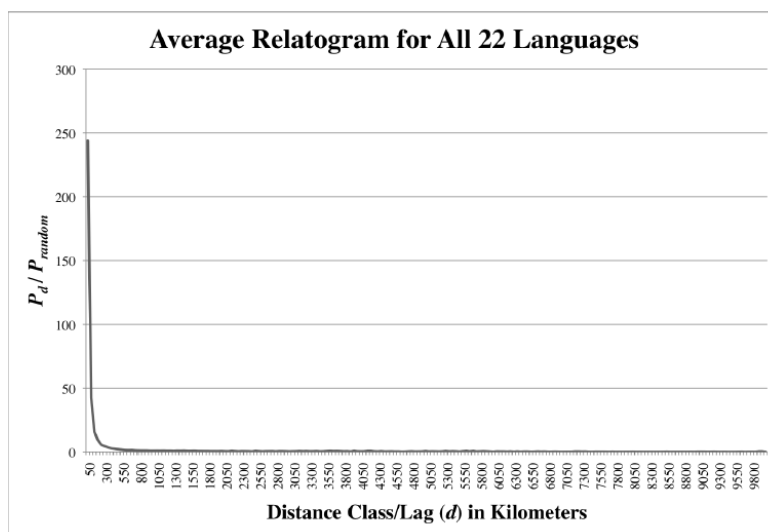
$$P_{relation}(A, B) = \frac{|E_{adjusted}|}{|V|^2 - |V|} = P_{random}$$

dove $|V|$ è il numero di articoli appartenenti a una certa edizione di Wikipedia e $|E_{adjusted}|$ è il numero dei *link* non duplicati⁹⁶.

Rapportare la probabilità che si verifichi un *link* tra due entità poste a una certa distanza alla probabilità che si verifichi un *link* tra due entità random, permette di calcolare la probabilità con cui si verifichi un *link* dall'articolo random spaziale A all'articolo random spaziale B [Tobler, 2004]:

$$\frac{P_d}{P_{random}}$$

⁹⁶ Si è in presenza di *link* duplicati quando i wikipedian aggiungono due o più link di collegamento in ogni articolo che genera una relazione verso l'altro.



(Figura 6.9. Relatogramma che presenta sull'asse y le probabilità che un link tra A e B possa verificarsi, data la distanza d; nell'asse x presenta le diverse classi di distanza. Più aumenta la distanza, più la probabilità che si stabilisca il link tende a 0. Fonte: Hecht e Moxley [2009]).

Come è possibile notare dalla figura 6.9, se gli articoli spaziali A e B si trovano a una distanza d che si aggira intorno ai 50 km, le volte in cui è probabile che ricorra una relazione si aggirano intorno alle 245 volte. Quando la distanza aumenta, la correlazione spaziale positiva tende allo zero. In altre parole, l'ipotesi iniziale è verificata:

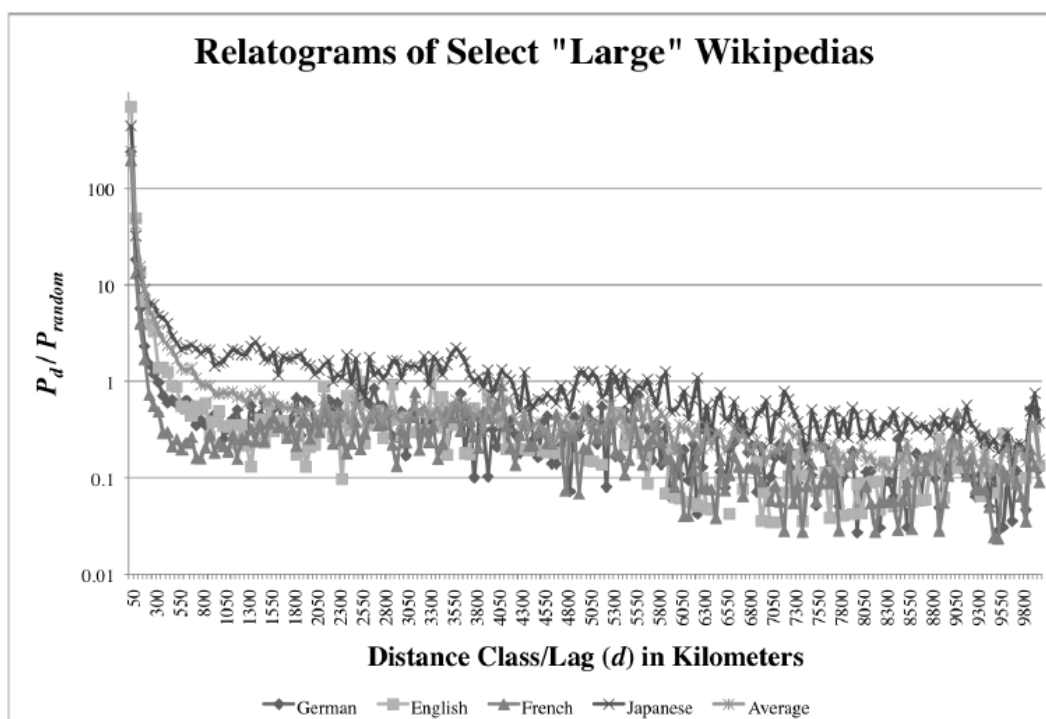
$$P_{\text{relation}}(A,B) > P_{\text{relation}}(A,C)$$

Se $d(A,B) < d(A,C)$.

Inoltre, la probabilità che A abbia una relazione con C non è mai uguale a zero.

Esiste una corrispondenza quasi manifesta tra P_d e il Coefficiente di Clusterizzazione [Padgett, 1999] utilizzato nella *Social Network Analysis*. Si ricorda, qui, che dato il nodo n con k vicini, il Coefficiente di Clusterizzazione di n misura il rapporto tra il numero di connessioni esistenti tra i k vicini di n ed il numero totale di connessioni possibili tra di loro:

$$k(k-1)/2.$$



(Figura 6.10. Clusterizzazione articoli spaziali in lingua inglese, tedesca, francese e giapponese. Fonte: Hecht e Moxley [2009])

Si noti dal relatogramma raffigurato in Figura 6.10. quanto siano clusterizzati gli articoli spaziali prodotti nelle lingue inglese, francese e tedesco e quanto tale clusterizzazione sia meno accentuata tra gli articoli prodotti in tali lingue e quelli corrispondenti elaborati in lingua giapponese. Il valore del coefficiente di clusterizzazione è più grande ad una classe di distanza di 50 km e, quindi, per tutti quei casi in cui gli articoli sono stati prodotti in lingua diversa da *wikipedian* vicini geograficamente.

Ragioni culturali e linguistiche influenzano certamente la forma di tale relazione. In particolare, gli individui acquisiscono con più facilità le specifiche norme linguistiche e le strutture grammaticali di alcune lingue se queste appartengono allo stesso ceppo a cui appartiene la propria (ad esempio, le lingue inglese e tedesco appartengono al ceppo germanico). Di conseguenza, essi manifestano una spiccata propensione a collegare i propri articoli a quelli corrispondenti prodotti dai *wikipedian* culturalmente e linguisticamente più vicini.

6.5. Funzione di interazione spaziale e diagrammi di Voronoi: oltre la distanza

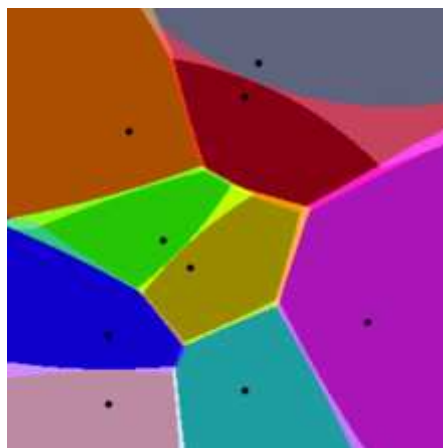
La logica funzionalista, intesa quale criterio di strutturazione delle attività produttive e urbane, nasce nel periodo di forte espansione economica degli anni Cinquanta e Sessanta. «L'industria e la città divengono dei sistemi altamente complessi, in cui agiscono attori numerosi ed eterogenei, che svolgono ruoli specializzati e che danno luogo ad una tipica articolazione secondo sottosistemi funzionali. (...). Tale articolazione può essere interpretata unicamente come l'effetto del processo di divisione del lavoro, che impone che ciascuna parte del sistema si specializzi nel raggiungimento di obiettivi relativamente autonomi, anche se subordinati ad un quadro più generale di esigenze della totalità. Tuttavia, ciò che trasforma una logica semplicemente funzionale in una logica funzionalistica è la pretesa del principio di articolazione funzionale del sistema a valere come unico paradigma della razionalità, come unico criterio in base al quale valutare l'efficienza di una struttura organizzativa e, dunque, la sua capacità di soddisfare esigenze e bisogni sociali» [Mela, 1992, 143]. Insomma, l'attore sociale e le sue azioni possono essere comprese solo a partire dal suo radicamento in una struttura che non può essere generalizzata ma subisce l'azione di fattori storici che vanno di volta in volta individuati [Martinotti, 2011, 124]. Lo spazio fisico, insomma, ha caratteristiche socio-economiche e politiche ben definite, ha un'identità riconosciuta nel tempo. La città non è una macchina costruita secondo criteri esclusivamente razionali. È anche un prodotto della storia. Ciascuna epoca storica non deposita i propri manufatti su quelli precedenti. Lo stesso vale per le proprie strutture socio-economiche, i propri sistemi simbolici [Bonafous e Puel, 1983]. Nonostante l'immagine ideale della città sembri riflettersi in una logica organizzativa in cui i soggetti sociali interagiscono in vista di obiettivi che consistono nell'aumento della produttività, la strutturazione sociale dello spazio cede inevitabilmente il passo a un sistema urbano complesso costituito da una forte eterogeneità tra le parti della città, determinata dal processo di complessificazione dei ruoli e delle attività e dalle differenze socio-culturali tra i soggetti.

La questione è stata oggetto di un lungo dibattito che non interessa il presente lavoro. Tuttavia, occorre sottolineare che, considerata la valenza della crisi in atto determinata da una messa in discussione del principio durkhemiano di divisione del lavoro tra sottosistemi funzionali, è opportuno applicare tecniche di analisi dei fenomeni situati spazialmente che permettano di analizzare lo spazio come l'insieme delle relazioni sociali di un individuo o di un'organizzazione che assume forme diverse in base ai legami esistenti al suo interno.

La struttura delle reti contribuisce a trasformare lo spazio in luogo, così come esso stesso ha contribuito a formare quella specifica struttura relazionale [Granovetter, 1973].

Una delle tecniche ampiamente utilizzate nell'ambito della *GIS analysis* e in cui lo spazio è considerato come un'astrazione geometrica, analizzato come distanza tra nodi, ampiezza, forma della relazione è il Diagramma di Voronoi [Gold, 2006], dal nome del russo Georgij Voronoi.

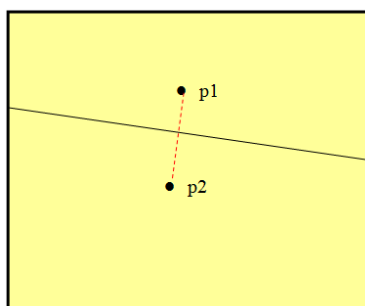
Tale Diagramma è costituito da regioni in cui ognuna contiene un punto del *set* dato (ad esempio, un set di indirizzi di aziende confiscate alla mafia o delle abitazioni di un villaggio in cui si registrano casi di febbre *dengue*).



(Figura 6.11. Diagrammi di Voronoi)

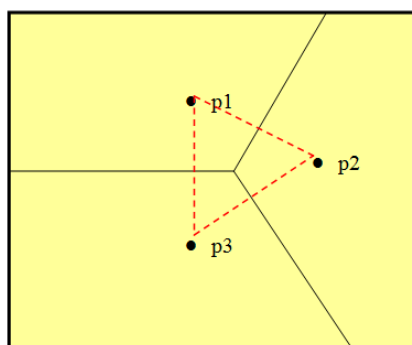
Dati i punti nel piano p_1 e p_2 , la bisecante della congiungente i due punti che divide il piano in due semipiani, è costituita da punti equidistanti da p_1 e p_2

[*Ibidem*, 4]. Significa che ogni punto localizzato nel semipiano intorno a p_1 è necessariamente più vicino a p_2 che a p_1 .



(Figura 6.12. Diagramma di Voronoi. Fonte: <http://xlr8r.info/mPower/examples.html>)

Dati tre punti nel piano, le tre bisecanti delle congiungenti $C_1(p_1,p_2)$, $C_2(p_1,p_3)$, $C_3(p_2,p_3)$, genereranno un punto nel piano (*punto di Voronoi*), dato dalla loro intersezione. Tale punto sarà equidistante dai tre punti p_1 , p_2 , p_3 attorno ai quali si costruiscono le regioni o *celle di Voronoi*. Ogni punto sul piano appartiene a un solo poligono, eccezione fatta per i confini che sono condivisi tra i poligoni tra loro adiacenti [*Ibidem*, 6].



(Figura 6.13. Diagramma di Voronoi. Fonte: <http://xlr8r.info/mPower/examples.html>)

Tale modello formalizzato di ripartizione del territorio [Boffi, 2004, 80] è implementabile attraverso i software GIS che elaborano il diagramma identificando l'area attorno ai punti dati. I Diagrammi di Voronoi aiutano a rispondere a domande come: «Quale abitazione è più vicina ad una considerata come focolaio *dengue*?». La tassellazione ripartisce e decompone territorialmente

lo spazio metrico, associando un peso ad ogni sito (dato, ad esempio, dal numero di casi di febbre *dengue* in una stessa abitazione). La metrica utilizzata per il calcolo delle distanze viene influenzata in modo additivo da tale peso. «La tassellazione di Voronoi è un particolare tipo di decomposizione dello spazio metrico determinato dalle distanze rispetto a un insieme discreto di elementi dello spazio. (...). Il modello genera una ripartizione spaziale (*area di vicinato*) determinando una collezione di regioni che ripartiscono il piano in maniera continua. (...). Tutti i punti all'interno della regione assumono lo stesso peso unitario. Al contrario, fuori dal poligono, tutti gli altri punti assumono peso nullo non influenzando, di conseguenza, i vicini» [Paolillo, 2011, 578-585].

Le formalizzazioni di Voronoi si sviluppano a partire dai principi della *funzione di interazione spaziale* [Butts, 2003] secondo i quali lo spazio fisico è costituito da nodi appartenenti a *network*, ciascuno con forme di radicamento spaziale diverse. La *funzione di interazione spaziale* permette di pensare per processi, esplorare come le forme fisiche dello spazio siano implicate nella formazione dei fenomeni, prevederne la direzione e la dimensione misurando le proprietà strutturali del contesto spaziale in cui hanno luogo, senza cadere nella trappola di ipostatizzare deterministicamente una relazione di causa effetto tra forma fisica e fenomeno. Secondo i suoi teorici [Wilson, 1970; Tobler, 1976; Brunson *et al.*, 1999], i rapporti tra le attività che si manifestano in un certo territorio sembrano organizzarsi seguendo le regole del campo di gravitazione [Newton, 1965] che, per definizione, è sensibile sia alle dimensioni delle attività localizzate sul territorio e sia alla loro distanza relativa⁹⁷. Nel caso dei Diagrammi di Voronoi, i corpi sono i punti che compongono il set dato, con coordinate X e Y, la loro massa è data dalla frequenza delle attività presenti nel punto noto. Specificatamente, in un modello di gravitazione, l'intensità di interazione è data da:

$$T_{ab} = K \cdot (M_a \cdot M_b) / d_{ab}^2$$

⁹⁷ Secondo la teoria di Newton, due corpi, rispettivamente di massa m_1 ed m_2 , si attraggono con una forza d'intensità direttamente proporzionale al prodotto delle masse e inversamente proporzionale al quadrato della distanza che li separa.

dove T è l'intensità d'interazione, M la dimensione dei due oggetti a e b , d è la distanza e K è la costante che dipende dall'unità di misura adottata. Ciò significa che il calcolo di tale intensità si emancipa dalla sola distanza fisica, avvicinandosi sempre più al concetto in base al quale la rete sempre più complessa fa la prossimità.

È chiaro come lo spazio assuma un significato più concreto e fisico quando si analizza la collocazione geografica dei nodi. I Diagrammi di Voronoi rappresentano la mobilità e l'immobilità geografica di una struttura relazionale oltre che il livello di *embeddedness* relazionale e strutturale. «L'attore sociale e le sue azioni sociali possono essere comprese solo a partire dal suo radicamento nella struttura delle reti sociali che lo circondano» [Granovetter, 1991, 59 in Nuvolati, 2011, 210].

Il concetto di spazio non è quasi mai tematizzato in modo esplicito dai ricercatori SNA, né considerato una questione teorica da approfondire. Tuttavia, finiscono sempre per discutere di uno spazio declinato come spazio sociale che plasma e definisce lo spazio fisico [Degenne e Forsé, 1999]. Quando analizzano le relazioni sociali e la loro struttura, guardano alla loro costruzione storica, alla loro evoluzione in una prospettiva dinamica; verificano come si siano cristallizzate le loro specifiche configurazioni e quali siano gli attori sociali e gli eventi chiave che le costituiscono. Analizzare le reti sociali implica la necessità di effettuare continue incursioni in campo urbano e richiamare costantemente i concetti di spazio e confine.

Perciò, da un lato, l'analista GIS-SNA avrà a che fare con uno spazio che assume forme differenti in base al tipo di legami esistenti al suo interno, uno spazio dai confini labili e mutevoli nel tempo, un'astrazione geometrica la cui forma dipende dalla distanza tra i nodi, ampiezza, forma, direzione della relazione, densità. Dall'altro lato, tale spazio assumerà un significato esplicitamente fisico quando dall'analisi formale il ricercatore passerà all'analisi dei nodi che compongono la rete e all'individuazione della loro posizione *spatially embedded*.

BIBLIOGRAFIA

- Abbott, A. (1997), *Of time and space: the contemporary relevance of the Chicago School*. In 'Social Forces', 75, pp. 1149-1182;
- Abbott, A. (2001). *Time Matters: On Theory and Method*, Orientale Institute Publication, University of Chicago Press;
- Abel, D.J., and D.M. Mark. (1990). *A comparative analysis of some 2-dimensional orderings*. International Journal of Geographical Information Systems 4(1): 21–31;
- Aberer Karl, Philippe Cudre-Mauroux, and Manfred Hauswirth. (2003). *The chatty web: Emergent semantics through gossiping*. In 12th World Wide Web Conference;
- Adams, J., Faust, K., Lovasi, G.S. (2012), *Special issue in the journal Social Network*. Volume 34. Issue 1, p. 2;
- Adar, E. and Huberman, B.A. (2000). 'Free Riding on Gnutella', *First Monday*, 5(10-2), Available at <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/792/701>;
- Addams, J. (1895). *Hull-house Maps and Papers. A presentation of Nationalities and Wages in a Congested District of Chicago*. New York: Library of Economics and Politics;
- Agrawal Parag, Omar Benjelloun, Anish Das Sarma, Chris Hayworth, Shubha Nabar, Tomoe Sugihara, and Jennifer Widom. (2006). Trio: A system for data, uncertainty, and lineage. In *Proceedings of the VLDB Conference*, pages 1151–1154;
- Agustoni, A. (1997), I luoghi urbani e il loro significato, in Guidicini, P. e Sgroi, E. (a cura di), *Valori, Territori, Ambiente*, FrancoAngeli, Milano;
- Al-Kodmany, K. (2002), GIS and the artist: shaping the image of a neighborhood through participatory environmental design, in Craig, W. Harris, T.e Weiner, D., *Community Participation and Geographic Information Systems*, London, Taylor e Francis, pp. 321-329;

Alexander, K. (2008), *Rdf in json*. In Proceedings of the 4th Workshop on Scripting for the Semantic Web, 20;

Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J. (2009), *Describing linked datasets*. In Proceedings of the WWW2009 Workshop on Linked Data on the Web, 48;

Alvisi, C., Ferrari, S., Nardini, E., Vaccari, A. (2013), *Accountability ed Open Data ambientali: gli Enti Locali alla sfida dell'apertura dei dati ambientali*. in Smart City Exhibition, La smart city al servizio del cittadino, ForumPA;

Allen, P. (1993), *Storia della cartografia. La rappresentazione del mondo nei più importanti atlanti geografici di tutte le epoche*, Mondadori, Milano;

Amaduzzi, S. (2011), *Geomarketing*, EPC Editore, Roma;

Amrhein, C. G. (1995), *Searching for the elusive aggregation effect: evidence from statistical simulation*. In 'Environment and Planning', A27, pp. 105-119;

Anderson, B.S., Butts, C.T., Carley, K.M. (1999), *The interaction of size and density with graph-level indices*, Social Networks, 21(3), 239-67;

Andrein, A., Clemente, P. (a cura di), (2007), *I Custodi delle Voci, Archivi orali in toscana: primo censimento*, in: "Toscana Beni Culturali", n. 8, Regione Toscana, Firenze;

Anselin, L. (1995), *Local indicators of spatial association*, in *Local Geographical Analysis* 27: 93–115;

Anselin, L. (1999), *Interactive techniques and exploratory spatial data analysis*. In P.A. Longley;

Anselin, L., Rey, S.J. (2010), *Perspectives on Spatial Data Analysis*, Berlin, Springer;

Aral, S., Muchnik, A. (2009), *Distinguishing influence-based contagion from homophily-driven diffusion in dynamic network*. Edited by Matthew O. Jackson, Stanford University, Stanford, CA;

Arbia, G., Benedetti, R., Espa, G. (1996), *Effects of the MUAP on image classification*, *Geographical Systems* 1. pages 123-141;

Atkinson P., Martin D. (2000), *GIS and Geocomputation*. London: Taylor and Francis;

Atzeni P., Ceri S., Paraboschi S., Torlone R. (2003). *Basi di dati. Concetti, Linguaggi e architetture*. McGraw-Hill;

Bailey, T.C., Gatrell, A.C. (1996), *Interactive Spatial Data Analysis*, Harlow, Longman;

Bangasser, P.E. (2000), *The ILO and the informal sector: an institutional history*, Geneva, International Labour Office;

Banfield, E.C. (1976), *Le basi morali di una società arretrata*, Il Mulino, Bologna;

Barbera, F. (2004), *Meccanismi sociali. Elementi di sociologia analitica*. il Mulino, Bologna;

Barthelemy, M. (2011), *Spatial Networks*. In ‘Physics Report’, 499, pp. 1-101;

Becker, C., Bizer, C. (2009), *Exploring the geospatial semantic web with dbpedia mobile*. *Journal of Web Semantics: Science, Services and Agents on the WorldWideWeb*, 7:278–286, 2009. DOI: 10.1016/j.websem.2009.09.004 86, 92;

Becker, C., Bizer, C., Erdmann, M., Greaves, M. (2010), *Extending smw + with a linked data integration framework*. In *Proceedings of the ISWC 2010 Posters & DemonstrationsTrack*, 93;

Beckett D.. (2004), *RDF/XML Syntax Specification* , W3C Recommendation. <http://www.w3.org/TR/rdf-syntax-grammar/>, 15, 18;

Belleau, F., Nolin, M., Tourigny, N., Rigault, P., Morissette, J. (2008), *Bio2rdf: Towards a mashupto build bioinformatics knowledge systems*. *Journal of Biomedical Informatics*, 41(5):706–16, DOI: 10.1016/j.jbi.2008.03.004 37;

Belussi, A. (2006), *Linguaggi di interrogazione per basi di dati geografiche: geo-algebra relazionale*, Dispense del corso Sistemi Informativi Geografici, Facoltà Scienze matematiche Fisiche e Naturali, Università di Verona;

Benjelloun, O., Das Sarma, A., Halevy, A. Widom, J. (2006), *Uldbs: Databases with uncertainty and lineage*. In VLDB;

Bergman, M.(2010), *What is a reference concept?* <http://www.mkbergman.com/938/whatis-a-reference-concept/>, 2010. 65

Berners Lee, T., Fielding, R., Masinter, L. (1998), *RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax*. <http://www.isi.edu/in-notes/rfc2396.txt>, August 1998, 7;

Berners Lee, T. (2001), *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*. Milano, Feltrinelli;

Berners Lee, T., Hendler, J., Lassilia, O. (2001), *The semantic web*. *Scientific American*, 284(5):34–44, Mai. DOI: 10.1038/scientificamerican0501-34 5;

Berners Lee, T. (1998), *Cool uris don't change*. <http://www.w3.org/Provider/Style/URI.1998>. 43

Berners Lee, T.. (2006), *Linked Data*, Design Issues; <http://www.w3.org/DesignIssues/LinkedData.html>. 7, 26, 82

Berners Lee, T. *et al.* (2006), *Tabulator: Exploring and analyzing linked data on the semantic web*. In Proceedings of the 3rd International Semantic Web User Interaction Workshop. <http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf>;

Berners Lee, T., Kagal L. (2008), *The fractal nature of the semantic web*. *AI Magazine*, Vol 29, No 3, 2008. 24, 62, 101, 107;

Berners Lee, T. (2009), *Putting government data online*, <http://www.w3.org/DesignIssues/GovData.html>, 2009. 36

Berners Lee, T. (2007), *Giant global graph*. <http://dig.csail.mit.edu/breadcrumbs/node/215>, (2007). 16, 29

Berners Lee, T. (2010), *Long live the web: A call for continued open standards and neutrality*. *Scientific American*, 32, 110;

Berrueta, D., Phipps J. (2008), *Best Practice Recipes for Publishing RDF Vocabularies.* W3C Working Group Note. <http://www.w3.org/TR/swbp-vocab-pub>;

Bertin J. (1967), *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Paris, Gauthiers-Villars & Mouton. Traduzione inglese: "*Semiology of graphics: diagrams, networks, maps*". (In particolare si veda il capitolo "*The properties of the graphic system*", p. 41-49);

Bertin J. (1977), *La graphique et le traitement graphique de l'information*. Flammarion, Paris;

Bezoari, Monti, Selvini. (2002), *Topografia generale con elementi di geodesia*, UTET;

Biallo, G. (eds.) (2013), *Dati Geografici Aperti: istruzioni per l'uso*. Associazione OpenGeoData Italia, Roma;

Bilke, A. Naumann, F. (2005), *Schema matching using duplicates*. In Proceedings of the International Conference on Data Engineering, 2005. DOI: 10.1109/ICDE.2005.126 102;

Biolcati-Rinaldi, F. e Vezzoni, C. (2013), *L'analisi secondaria nella ricerca sociale*, Il Mulino, Itinerari, Bologna;

Birardi (1988), *Corso di geodesia, topografia e fotogrammetria*, IGM;

Birbeck, M. (2009), *Rdfa and linked data in uk government web-sites*. *Nodalities Magazine*, 7, 36;

Biron, P. Ashok, M. (2004), *Xml schema part 2: Datatypes second edition - w3c recommendation*. <http://www.w3.org/TR/xmlschema-2/>, 2004. 16;

Bishop, W. (2008), *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*, Houghton Mifflin Harcourt;

Bizer, C. (2010), *Pay-as-you-go data integration on the public web of linked data*. Invited talk at the 3rd Future Internet Symposium 2010; <http://www.wiwiss.fuberlin.de/en/institute/pwo/bizer/research/publications/%Bizer-FIS2010-Pay-As-You-Go-Talk.pdf>, 2010. 107;

Bizer C., Schultz, A. (2009), *The berlin sparql benchmark*. *International Journal on SemanticWeb and Information Systems*, 5(2):1–24, 105;

Bizer C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R., Eckstein, R. (2005), *The impact of semantic web technologies on job recruitment processes*. In Proceedings of the 7. Internationale Tagung Wirtschaftsinformatik (WI2005), 2005. DOI: 10.1007/3-7908-1624-8_72 36;

Bizer, C. (2007), *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*. PhD thesis, Freie University at Berlin;

Bizer C., Heath, T., Berners-Lee, T. (2009), *Linked Data—The Story So Far*, Int'l. J. Semantic Web & Information Systems;

Bizer C., Cyganiak, R., Heath T.(2007), *How to Publish Linked Data on the Web*. <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial>;

Bizer C., Cyganiak, R. (2009), *Quality-driven information filtering using the wiqua policy framework*. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7(1):1–10, DOI: 10.1016/j.websem.2008.02.005 103, 104;

Bizer, C., Cyganiak, R., Gauss, T. (2007), *The rdf book mashup: From web apis to a web of data*. In *Proceedings of the Workshop on Scripting for the Semantic Web*, 38, 70;

Bizer, C., Schultz, A. (2010), *The r2r framework: Publishing and discovering mappings on the web*. In *Proceedings of the 1st International Workshop on Consuming Linked Data*, 25, 102;

Blacksher, F., Lovasi, G.S. (2011), *Place-focused physical activity research, human agency and social justice in public health: taking agency seriously in studies of the built environment*. In 'Health & Place';

Blanford, J., Mitra, P. (2011), *Geo-Twitter Analytics: Applications in Crisis Management*. *Proceedings of the 25th International Cartographic Conference*, Paris, France;

Bleiholder, J., Naumann, F. (2008), *Data fusion*. *ACM Computing Surveys*, 41(1):1–41, 2008. DOI: 10.1145/1456650.1456651 104;

Bobrowski S. (1995), *La grande guida a Oracle 7*. Jackson;

Boero, R. e Squazzoni, F. (2005), *Towards an Agent-Based Computational Sociology. Good reasons to Strengthen a Cross-Fertilization between Complexity and Sociology*. In 'Stoneham P.M., *Advances in Sociology Research II*, Nova Science Publishers Inc., New York, USA, pp. 103-133;

Boffi, M. (2004), *Scienza dell'Informazione Geografica*, Zanichelli;

Bonnafous, A., Puel, H. (1983), *Physionomie de la ville*, Les Editions ouvrière, Paris;

Booth C. (1892–1897), *Life and Labour of the People in London* (9 Volumes, Plus Maps under Separate Cover). London: Macmillan and Co.;

- Bossard, J.H.S. (1932), *Residential propinquità as a factor in marriage selection*, *American Journal of Sociology*, 38, 219-44;
- Boudon, R., Lazarsfeld, P. F. (1969), *L'analisi empirica nelle scienze sociali*. (vol. 1), Il Mulino, Bologna;
- Boudon, R. (1974), *Metodologia della ricerca sociologica*, il Mulino, Bologna;
- Boudon, R. (2003), *Beyond rational choice theory*. In 'Annual Review of Sociology' 29, pp. 1-21;
- Brandes, U. e Erlebach, T. (2005), *Network Analysis: Methodological Foundations*, Springer-Verlag, Berlin;
- Brodeur, J., Bedard, Y. Proulx, M.J., *Modelling geospatial application databases using UML-based repositories aligned with international standards in geomatics*, 8th ACM Symposium on GIS 11/00 Washington, DC, USA;
- Breslin, J., Andreas, H., Uldis, B., Decker, S. (2005), *Towards semantically interlinked online communities*. In *Proceedings of the 2nd European Semantic Web Conference*, Heraklion, Greece, 2005. DOI: 10.1007/11431053_34 54;
- Brickley, D., Guha, R. V (2004), *RDF Vocabulary Description Language 1.0: RDF Schema - W3C Recommendation*. <http://www.w3.org/TR/rdf-schema/>, 17, 24, 56;
- Brown, M.C. (2006), *Hacking Google Maps and Google Earth*. New York: Wiley
- Brownstein, J.S., Freifeld, C.C., Reis, B. Y., Mandl, K.D. (2008), *Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project*. *PLoS Med*, 5(7): 1019-1024;
- Brunsdon, C., Fotheringham, A.S., Charlton, M. (1999), *Somenotesonparametric significance testsfor geographically weighted regression*. *Journal of Regional Science* 39(3): 497–524;
- Buneman P., Sanjeev. K., Wang C.T. (2001), *Why and where: A characterization of data provenance*. In ICDT, volume 1973 of Lecture Notes in Computer Science;
- Butler, D. (2006), *Virtual globes: the web-wide world*. *Nature* 439: 776–778;
- Butts, C.T. (2003), *Network inference, error, and informant (in)accuracy: a Bayesian approach*, *Social Network*, 25(2): 103-40;

Butts, C.T., Acton, R.M., Marcum, C. (2011), *Interorganizational collaboration in the Hurricane Katrina response*, Journal of Social Structure, Forthcoming;

Calkins, H.W. (1996), *Entity-Relationship modeling of spatial data for geographic information systems*, International Journal of Geographical Information Systems;

Cannavò, L. e Frudà, L. (2009), *Ricerca sociale. Tecniche speciali di rilevazione, trattamento e analisi*, Carocci editore, Roma;

Carrington, P.J., Scott, J., Wasserman, S. (2005), *Models and Methods in social network analysis*, Cambridge University Press, Cambridge;

Carneiro, H.A., Mylonakis, E. (2009), *Google Trends: A Web-based Tool for Real-Time Surveillance of Disease Outbreaks*, *Clinical Infectious Diseases*, 49(10):1557-1564;

Castino M., Roletto, E. (1991), *Statistica Applicata*, Piccin, Padova;

Charmaz, K. (1995), *Grounded Theory*, in *Rethinking Methods in Psychology*, a cura di J.A. Smith, R. Harré e L. Van Langenhove, Sage, London;

Cheshire, James and Batty, Michael and Reades, Jonathan and Longley, Paul and Manley, E., Milton, R. (2013), *CyberGIS for Analyzing Urban Data*. In *CyberGIS: Fostering a New Wave of Geospatial Discovery and Innovation*. Springer-Verlag. Talisman node, University College London. Disponibile online sul sito <http://eprints.ncrm.ac.uk/3159/>;

Chrisman, N. (1998), *Rethinking Levels of Measurement for Cartography*, in *Cartography and Geographic Information Science*, vol. 25 (4), p. 231-242;

Christaller, W. (1972), *Come ho scoperto la teoria dei luoghi centrali: un rapporto circa l'origine dei luoghi centrali*. In inglese, PW e RC Mayfield, a cura di *Spazio Uomo e Ambiente*. Oxford University. Press, 1972, pp.601-610;

Cicognani E. (2002), *Psicologia sociale e ricerca qualitativa*, Roma, Carocci;

Cicognani, E. (2002), *L'approccio qualitativo della Grounded theory in psicologia sociale: Potenzialità, ambiti, applicazioni e limiti*, in *Metodi qualitativi in psicologia sociale*, a cura di B. Mazzara, Roma, Carocci, pp. 43-59;

- Clarke, K.C., Hoppen, S., Gaydos, L. (1997), *A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area*. *Environment and Planning B: Planning and Design* 24(2): 247–261;
- Clauset, A., Shalizi, C.R. and Newman, M.E.J. (2009), *Power-law Distributions in Empirical Data*, *SIAM Review*, 51(4): 661-703;
- Clodoveu, A.D, Fonseca, F., Camara, G. (2011), *Environmental Sustainability: The Role of Geographic Information Science and Spatial Data Infrastructures in the Integration of People and Nature*, in *The Sage Handbook of GIS and Society*, Sage;
- Conte, R., Edmonds, B., Moss, S., Sawyer, R.K. (2001), *Sociology and Social Theory in Agent Based Social Simulation: A Symposium*, in *Computational and Mathematical Organization Theory*, n. 7, 183-205
- Corbett, J.M., Keller, C.P. (2005), *An analytical framework to examine empowerment associated with participatory geographic information systems (PGIS)*, in *Cartographica*, 40, pp. 91-102;
- Corbetta, P., Gasperoni, G., Pisati, M. (2001), *Statistica per la ricerca sociale*, Il Mulino, Bologna;
- Corley, C.D., Cook, D.J., Mikler, A.R. and Singh, K.P. (2010), *Text and Structural Data Mining of Influenza Mentions in Web and Social Media*, *International Journal of Environmental Research and Public Health*, 7(2): 596-615;
- Crampton, J. (2009), *Being ontological: response to Poststructuralism and GIS: Is there a disconnect?*, *Environment and Planning, Society and Space*, 27(4), 603-608;
- Crang, M. (2005), *Qualitative methods: there is nothing outside the text?*, *Progress in Human Geography*, 29 (2);
- Creswell, J.W. (2003), *Qualitative Inquiry and Research Design*, Sage, Thousand Oaks (CA);
- Crooks, A., Croitoru, A., Stefanidis, A., e Radzikowski, J. (2011), *#Earthquake: Twitter as a Distributed Sensor System*, *Transactions in GIS*;

Culotta, A. (2010), *Towards Detecting Influenza Epidemics by Analyzing Twitter Messages, Proceedings of the First Workshop on Social Media Analytics*, Washington, DC, pp. 115-122;

Cyganiak, R., Hellmann, S. (2009), *Dbpedia. A crystallization point for the web of data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165. DOI: 10.1016/j.websem.2009.07.002 14, 33;

Dainelli, N., Bonechi, F., Spagnolo, M., Canessa, A. (2010), *Cartografia numerica*, Flaccovio, Palermo;

Darra, A. (2009), *A framework for the development and comparison of spatial socioeconomic units*, PhD thesis, National Technical University of Athens;

de Certeau, M. (2001), *L'invenzione del quotidiano*, trad. M. Baccianini, Edizioni Lavoro, Roma;

Date, C.J. (2003), *An Introduction to database systems*, Addison Wesley;

De Vera, D. (2005), *Mapping with communities in the Philippines*, paper presented at the Mapping for Change Conference, Nairobi, Kenya;

De Gregorio, E., Mosiello, F. (2004), *Tecniche di ricerca qualitativa e di analisi delle informazioni con Atlas.ti*, Edizioni Kappa, Roma;

Decker, H., Teniente, E., Urpì, T. (1999), *How to Tackle Schema Validation by View Updating "SQL-99 Complete, Really"*, R&D Books, 1999;

Degenne, A., Forsé, M. (1999), *Introducing Social Networks*, Sage, London;

Denzin, N.K., (1978), *The research act: a theoretical introduction to sociological methods*, New York, McGraw-Hill;

DiBiase, D. (1990), *Visualization in the earth sciences*, in *Earth and Mineral Sciences*, Bulletin of the College of Earth and Mineral Sciences, Vol. LIX, p. 13-18;

Donati, P. (1991), *Teoria relazionale della società*, Angeli, Milano;

Doreian, P., Batagelj, V., Ferlioj, A. (2005), *Generalized Blockmodeling*, Cambridge University Press, Cambridge;

Dorling, D. (1995), *A New Social Atlas of Britain*. Chichester, U.K.: Wiley;

Durkheim, È. (1893, trad. it. 1962), *La divisione del lavoro sociale*, Comunità, Milano.

- Durkheim. È., (1895, trad. it. 1964), *Le regole del metodo sociologico*, Comunità, Milano.
- Elster, J. (1993). *Come si studia la società. Una cassetta degli attrezzi per le scienze sociali*, il Mulino, Bologna;
- Elwood, S. (2006), *Beyond cooptation or resistance: urban spatial politics, community organizations, and GIS-based spatial narratives*, Annals of the Association of America Geographers, 96 (2);
- Eriksson, B., Barford, P., Sommers, J. and Nowak, R. (2010), A Learning-Based Approach for IP Geolocation, in Krishnamurthy, A. and Plattner, B. (eds.), *Passive and Active Measurement, Lecture Notes in Computer Science*, Volume 6032, Springer, Berlin, Germany, pp. 171-180;
- Erling, O., Mikhailov, I. (2007), *RDF support in the Virtuoso DBMS*. volume P-113 of GI-Edition, Lecture Notes in Informatics (LNI);
- ESRI, *Supported Open Geospatial Consortium*, Inc., and ISO/TC211 Standards, ESRI White Paper;
- Estes, J.E., Mooneyhan, W. (1994), *Of maps and myths. Photogrammetric Engineering and Remote Sensing* 60: 517-524;
- Field, K., O'Brien, J. (2010), *Cartoblography: Experiments in Using and Organising the Spatial Context of Micro-blogging*, Transactions in GIS 14, 5-23;
- Fiedler, R., Schuurman, N. e Hyndman, J. (2006), *Improving census-based socioeconomic GIS for public policy: recent immigrants, spatially concentrated poverty and housing need in Vancouver*, ACME, An International E-Journal for critical Geographies, 4 (1), 145-171;
- Fielding, N.G. e Lee, R.M. (1998), *Computer Analysis and Qualitative Research*, London, Sage;
- Fielding, N.G. (2010), *Virtual Fieldwork using Access Grid*. In 'Field Methods', 22(3), pp. 195-216;
- Fischer, M.M. e Getis, A. (2010), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Berlin, Springer;
- Florida, R. (2005), *Cities and the creative class*, Routledge, London - New York;

- Flowerdew, R., Manley, D. J., Sabel, C. E. (2008), *Neighbourhood effects on health: does it matter where you draw the boundaries?* In 'Social Science & Medicine', 66, pp. 1241-1255;
- Fotheringham, A.S., e O'Kelly, M.E. (1989), *Spatial Interaction Models: Formulations and Applications*, Dordrecht, Kluwer;
- Fotheringham, A.S., Charlton, M.E. e Brunson, C. (1998), *Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis*. Environment and Planning A 30(11): 1905–1927;
- Fotheringham, A.S., e Brunson, C. (1999), *Local forms of spatial analysis*, Geographical Analysis 31(4): 340–358;
- Fotheringham, A.S., Brunson, C. e Charlton, M. (2000), *Quantitative Geography: Perspectives on Spatial Data Analysis*, London: Sage;
- Fowler, M. (2004), *Guida rapida al linguaggio di modellazione standard*. Pearson Education Italia;
- Frank, O. (2005), Network sampling and model fitting, in Carrington, P.J., Scott, J., Wasserman, S., *Models and Methods in Social Network Analysis*, Cambridge University Press, Cambridge;
- Friedland, G. e Sommer, R. (2010), *Cybercasing the Joint: On the Privacy Implications of Geotagging, Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10)*, Washington, DC;
- Fritz, S., MacCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. e Obersteiner, M. (2009), *Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover*, Remote Sensing, 1 (3): 345-354;
- Gastner, M.T., Newman, M.E.J. (2004), *Diffusion-based method for producing density equalizing maps*, Proc Natl Acad Sci USA, 101: 7499-7504;
- Gelman, A., Price, P., Lin, C. (2000), *A method for quantifying artefacts in mapping methods illustrated by application to headbanging*. Statistics in Medicine, 19: 2309-20;
- Getis A., e Ord, J. K. (1992), *The analysis of spatial association by use of distance statistics*. Geographical Analysis 24: 189–206;

- Gettys, W.E. (1961), *Human ecology and social theory*, in G.A. Theodorson (a cura di), *Studies in human ecology*, Harper&Row, NY, 98-103;
- Giannelli, L. (2008), *Percezione mapuche del Bosque austral*, in: *Quaderni di Thule*, n. 8, pp: 415 - 420;
- Giddens, A. (1981), *A contemporary critique of Historical Materialism, vol. 1*, Power, Property and the state, Macmillan, London;
- Giddens, A. (1990), *La costituzione della società. Lineamenti di teoria della strutturazione* (1984), Edizioni di Comunità, Milano;
- Gilbert, E. W. 1958. Pioneer maps of health and disease in England. *Geographical Journal* 124:172–183;
- Gilbert, M.R. e Masucci, M. (2005), *Research directions for information and communication technology and society in geography*, *Geoforum*, 36, 277-279;
- Glaser e Strauss (1967), *The discovery of the grounded theory: strategies for qualitative research*, Aldine, New York;
- Goodchild, M.F. (1992), Analysis. In Abler, R.F., Marcus, M.G., e Olson, J.M. (eds.), *Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography*. New Brunswick, N. J.: Rutgers University Press, 138–162;
- Goodchild, M.F. (2008), Epilogue: Intelligent systems for GIScience: Where next? A GIScience perspective. In Agarwal, P. e Skupin, A., *Self-Organizing Maps: Applications in Geographic Information Science*. Chichester, Wiley, pp. 195–199;
- Goodchild, M.F. (1986), *Spatial Autocorrelation*. *Catmog* 47, Geo Books;
- Goodchild, M.F. (2007a), *Citizens as Sensors: The World of Volunteered Geography*, *GeoJournal*, 69(4): 211-221;
- Goodchild, M.F., Maguire, D.J., Rhind, D.W. (eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley, 253–266.
- Gold, C.M. (2006), *What is GIS and what is not?*, *Transaction in GIS*, 10(4).
Disponibile on line sul sito
http://www.voronoi.com/wiki/images/c/c8/What_is_gis.pdf;

- Goramasca, M. (2004), *Elementi di geomatica*, Associazione Italiana di telerilevamento;
- Grannis, R. (1998), *The importance of trivial streets: residential streets and residential segregation*, In *American Journal of Sociology*, 103, 1530-1564;
- Granovetter, M. (1973), La forza dei legami deboli. In *La forza dei legami deboli ed altri saggi*, Liguori, Napoli, pp. 115-146;
- Granovetter, M. (1991), Azione economica e struttura sociale. Il problema dell'embeddedness (1985), in Magatti, M. (a cura di), *Azione economica come azione sociale*, 49-81, FrancoAngeli, Milano;
- Griffith, D. (1987), *Spatial Autocorrelation: A Primer*, Resource Publications in Geography, Association of American Geographers;
- Guidicini, P. e Sgroi, E. (1997), *Valori, territorio, ambiente*, FrancoAngeli, Milano;
- Guidi A., Dordolò, D. (1996), *SQL*, McGraw-Hill;
- Güting, R. H. (1994), *An introduction to spatial database systems*, VLDB Journal, Vol. 3, n. 4;
- Hakim, C. (2000), *Research Design*, Routledge, London-New York;
- Handcock, M.S., Raftery, A.E., Tantrum, J.M. (2007), *Model based clustering for social network*, *Journal of the Royal Statistical Society, Series A*, 170, 301-54;
- Hatt, P. (1946), *The concept of natural area*, in *American Sociological Review*, XI, 423-427;
- Haynes, K.E., Fotheringham, A.S. (1984), *Gravity and Spatial Interaction Models*, Sage, Beverly Hills, CA;
- Hecht, E., Moxley, E. (2009), *Terabytes of Tobler: Evaluating the First Law in a Massive, Domain-Neutral Representation of World Knowledge*. *Spatial Information Theory Lecture Notes in Computer Science Volume 5756*, 2009, pp. 88-105;
- Hedström, P. (2005), *Anatomia del sociale. Sui principi della sociologia analitica*, Mondadori, Milano;

- Hedström, R. e Swedberg, R. (a cura di), (1998), *Social Mechanism: An Analytical Approach to Social Theory*, Cambridge University Press, Cambridge, 238-266;
- Heilbrun, J. (1987), *Economia urbana e ordine pubblico*, la terza edizione, New York, St Martins Pr;
- Hilbert, D. (1960), *Geometria intuitiva*, Boronghieri, Torino;
- Hipp, J. R., (2012), *Geographical variability and network structure*. In 'Social Networks', 34, pp. 82-100;
- Hipp, J.R., Perrin, A.J. (2009), *The simultaneous effect of social distance and physical distance on the formation of neighborhood tie*, City and Community, 8(1), 5-25;
- Hoff, P.D., Raftery, A.E., Handcock, M.S. (2002), *Latent space approaches to social network analysis*, Journal of the American statistical association, 97(460), 1090-98;
- Holland, P.W., Leinhardt, S. (1971), *Transitivity in structural models of small groups*, Comparative Group Studies, 2, 107-24;
- Holstein, J.A. e Gubrium, J.F. (1997), *Active Interviewing*, in Qualitative Research: Theory, Method and Practice, a cura di D. Silverman, London, Sage;
- Howe, J. (2006), *The Rise of Crowdsourcing*, Wired, 14.06: 161-165, Disponibile sul sito <http://www.wired.com/wired/archive/14.06/crowds.html>;
- Hudson-Smith, A. e Crooks, A.T. (2009), *The Renaissance of Geographic Information: Neogeography, Gaming and Second Life*, in Lin, H. and Batty, M. (eds.), *Virtual Geographic Environments*, Science Press, Beijing, PRC, pp. 25-36;
- Intesa GIS (2004), *Il modello concettuale GeoUML: inquadramento generale e introduzione all'uso*;
- Irwin, M.D., Hughes, H.L. (1992), *Centrality and the structure of urban interaction: measures, concepts and applications*, Social Forces, 71(1), 17-51;
- Jacobs, J. (1958), *Downtown is For People*, in W.H. White (a cura di), *The Exploding Metropolis*, Doubleday, New York, disponibile in lingua italiana all'indirizzo <http://mall.lampnet.org/article/articleview/12835/0/212/>;

- Jacques J. (2007), Spatial cluster analysis, in Wilson, J., Fotheringham, S. (a cura di), *The Handbook of Geographic Information Science*, Wiley-Blackwell, Chicester;
- Janelle, D.G. e Goodchild, M.F. (2011), Concepts, principles, tools, and challenges in spatially integrated social science. In T.L. Nyerges, H. Couclelis, and R. McMaster, editors, *The SAGE Handbook of GIS and Society*, pp. 27–45. Thousand Oaks, CA: SAGE;
- Janelle, D.G e Hodge, D.C. (2000), *Information, Place and Cyberspace: Issues in Accessibility*, National Center for Geographic Information & Analysis, Springer;
- Jin-Yi Cai, Venkatesan, T., Chakaravarthy, Kaushik, R. e Naughton, J.F. (2001), *On the Complexity of Join Predicates*. PODS;
- Jones, J.H., Handcock, M.S. (2003), *An assessment of preferential attachment as a mechanism for human sexual network formation*, Proceeding of royal society, Series B, 270, 1123-28;
- Jung, J. (2009), Computer-aided Qualitative GIS: A software-level Integration of Qualitative Research and GIS, in Cope M. e Elwood S., *Qualitative GIS*, Sage Publication Ltd, London;
- Kang, H. e Scott, D.M. (2007), *An integrated spatio-temporal GIS toolkit for exploring intra-household interactions*, 86th Annual Meeting of the Transportation Research Board: Compendium of Paper CD-Rom, Washington, DC, 21-25 January;
- Kaushik, R., Krishnamurthy, R., Naughton, J.F., Ramakrishnan, R. (2004), *On the Integration of Structure Indexes and Inverted Lists*, In Proc. of ACM SIGMOD, Paris, France, 779-790;
- Klathen, K. (2007), *Norway SDI: Building sustainability through strategic alliances*, in Współpraca i Koordynacja w Zakresie Geoinformacji, Waraw, PTIP;
- Kruglanski, A.W. e Jost, J.T. (2000), *Il costruzionismo sociale e la psicologia sociale sperimentale: storia delle divergenze e prospettive di riconciliazione*, in Rassegna di Psicologia, 17 (3), pp. 45-67;
- Krivitsky, P.N., Handcock, M.S. (2008), *Fitting latent cluster models of networks with latentnet*, Journal of Statistical software, 24(5);

- Kwan, M.P. e Knigge, L. (2006), *Guest editorial. Doing qualitative research using GIS: an oxymoronic endeavor?*, Environment and Planning A, 38 (11), 1999-2002;
- Kwan, M. (2007), *Affecting geospatial technologies*, The Professional Geographer, 59 (1), 22-34;
- La Cecla, F. (2000), *Perdersi. L'uomo senza ambiente*, Laterza, Roma-Bari;
- Labow (1997), *Some Further Steps in Narrative Analysis*, in Journal of Narrative and Life History, 7 (1-4);
- Lake, R. (1993), *Planning and applied geography: positivism, ethics, and geographic information systems*, Progress in Human Geography, 17 (3), 404-413;
- Lee, C., Moudon, A. V., Courbois, J. Y., (2005), *Built environment and behavior: spatial sampling using parcel data*. In 'Annals of Epidemiology', 16, pp. 387-394;
- Leitner, H., Elwood, S., Sheppard, E., McMaster, R. (2000), *Modes of GIS provision and their appropriateness for neighborhood organizations: examples from Minneapolis and St Paul*, Minnesota, The URISA Journal, 12(4), 43-56;
- Lewins, A. e Silver, C. (2004), *Choosing a CAQDAS package*, CAQDAS Networking Project, testo disponibile online su <http://caqdas.soc.surrey.ac.uk>;
- Logan, J. e Collver, A. (1983), *Residents' Perception of Suburban Community Differences*, American Sociological Review, 48, 3;
- Longley, P.A., Goodchild, M.F., Maguire, D.J. e Rhind, D.W. (2010), *Geographical Information Systems and Science* (3rd Edition), John Wiley & Sons, New York, NY;
- Luhman, N. (1973), *Teoria della società o tecnologia sociale*, ETAS Kompass, Milano;
- Lynd, R.S. e Lynd, H.M. (1970), *Middletown*, vol. I, Edizioni di Comunità, Milano;
- MacEachren, A.M. (1994), *Visualization in modern cartography*, Pergamon Press, Oxford;
- MacEachren, A.M. (1994), *Some truth with maps: a primer on symbolization and design*, Association of American Geographers, Washington;

- MacEachren, A.M., Robinson, A.C., Jaiswal, A., Pezanowski, S., Saveljev, A., Mathioudakis, M., e Koudas, N. (2010), *TwitterMonitor: Trend Detection over the Twitter Stream*, Proceedings of the International Conference on management of Data – SIGMOD'10, Indianapolis, pp. 1155-1158;
- Macintyre, S., Ellaway, A., Cummins, S. (2002), *Place effects on Health: how can we conceptualise, operationalise and measure them?* In 'Social Science & Medicine', 55, pp. 125-139;
- Mangabeira, W. (1995), Computer assistance, qualitative analysis and model building, in *Information technology for the social scientist*, a cura di R.M. Lee, London, UCL Press;
- Mannetti, L. (1998), (a cura di), *Strategie di ricerca in psicologia sociale*, Roma, Carocci;
- Mantovani, G. (2003), I metodi qualitativi in psicologia. Strumenti per una ricerca situata, in *Metodi qualitativi in psicologia sociale*, a cura di G. Mantovani e A. Spagnoli, Bologna, Il Mulino;
- Marradi A. (1987). *Concetti e metodo per la ricerca sociale*, Firenze, La Giuntina;
- Marsden, P. V. (2011), *Survey methods for network data*. In 'Scott, J. (ed.), *The Sage Handbook of Social Network Analysis*, Sage, Los Angeles, CA;
- Martin, D., Cockings, S. and Harfoot, A.(2012), Development of a geographical framework for Census workplace data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*;
- Martinotti, G. (1993), *Metropoli. La nuova morfologia sociale della città*, Il Mulino, Bologna;
- Martinotti, G. (2011), Robert E. Park: l'ecologia umana, in Nuvolati, G. (a cura di), *Lezioni di Sociologia urbana*, Il Mulino, Bologna;
- Marx, K. (1859), *Per la critica dell'economia politica*, Roma, Editori Riuniti;
- Massey, D. (1993), Power-Geometry and a Progressive Sense of Place, in J. Bird, Curtis, B., Putnam, T., Robertson, G., Tickner, L. (a cura di), *Mapping the Futures: Local Cultures, Global Change*, 59-69, Routledge, London;
- Mazzara B. (2002), *Metodi qualitativi in psicologia sociale*, Roma, Carocci;

McCleary, G.F.J. (1983), *An effective graphic vocabulary*. IEEE Computer Graphics & Applications 3, n.2 ,46-53;

McCoy, J. (2004), *ArcGIS 9, Geoprocessing in ArcGIS*, Environmental Systems Research Institute, Redlands California;

McMaster, R. e Harvey, F. (2010), *Geographic Information Science and Society*, in Bossler, J.D., Campbel, J.B. e McMaster, R., *A Manual of Geospatial Technology*, London , Taylor & Francis;

McPherson, M., Smith-Lovin, L., Cook. J.M. (2001), *Birds of a Feather. Homophily in Social Networks*. Source: Annual Review of Sociology, Vol. 27 415-444 Published by: Annual Reviews;

Mela, A. (1992), *La città come sistema di comunicazioni sociali*, Franco Angeli, Milano;

Mela, A., Davico, L., Conforti, L. (2000), *La città, una molte. Torino e le sue dimensioni spaziali*, Liguori, Napoli;

Mela, A. (2006), *Sociologia delle città*, Carocci, Roma;

Mendoza, M., Poblete, B. e Castillo, C. (2010), *Twitter Under Crisis: Can we trust what we RT?*, 1st Workshop on Social Media Analytics (SOMA '10), Washington, DC, USA;

Merton, R.K. (1949, trad. it. 2000), *Teoria e struttura sociale, 3 voll.*, Il Mulino, Bologna;

Mingo, I. (2007), *Le fonti statistiche nella ricerca sociale*, in (a cura di) Cannavò, L. e Frudà, L., *Ricerca sociale. Da progetto dell'indagine alla costruzione degli indici*, Carocci Editore, Roma;

Mitchell, A. (2005), *The ESRI Guide to GIS Analysis, Volume 2*. ESRI Press;

Monmonier, M. (1993), *Mapping it out: expository cartography for the humanities and social sciences*, University of Chicago Press;

Monmonier, M. (1996), *How to lie with maps*, University of Chicago Press. Disponibile gratuitamente sul sito <http://courses.washington.edu/war101/readings/Monmonier.pdf>;

Monti, C., Pinto, L. (2002), *Trattamento dei dati topografici e cartografici*. Libreria Clup, Milano;

Montjoye, Y., Hidalgo, C., Verleysen, M., Blondel, V. (2012), *Unique in the Crowd: The privacy bound of human mobility*, Scientific Reports 3, Disponibile online sul sito <http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html>;

Moran, P.A.P. (1950), *Notes on Continuous Stochastic Phenomena*, Institute of Statistics, Oxford University;

Morenoff, J.D., Sampson, R.J., Raudenbush, S. (2001), *Neighborhood Inequality, Collective efficacy, and the spatial dynamics of urban violence*, Criminology, 39, 517-570;

Morrill, R.L., Pitts, F.R. (1963), *Marriage, migration and the mean information field: a study in uniqueness and generality*, Annals of the Association of American Geographers, 57, 401-422;

Morris, M. (1991), *A log-linear modeling framework for selective mixing*, Mathematical Biosciences, 107, 349-77;

Morrison, J.L. (1984), *Applied cartographic communication: map symbolization for atlases*, in Cartographica 21, n.1 , pp. 44-84;

Mutti, A. (1996), *Reti sociali: tra metafore e programmi teorici*, in Rassegna italiana di Sociologia, 1;

National Research Council, (1993), *Toward a Coordinated Spatial Data Infrastructure for the Nation*, Washington, DC, National Academies Press;

National Research Council, (2007), *Successful Response Starts with a Map: Improving Geospatial Culture*, Washington, DC, National Academies Press;

Newman, M. E. J. (2002), *Assortative mixing in networks*. Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501;

Newman, M. E. J. (2003), *Mixing patterns in networks*, Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120 and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501;

Newman, M.E.J. (2005), *Power Laws, Pareto Distributions and Zipf's Law*, Contemporary Physics, 46(5);

- Newton, I. (1965), *Principi matematici della filosofia naturale*, vol. Classici della scienza, Torino Utet;
- Norberg-Schulz, C. (1979), *Genius loci. Paesaggio, ambiente, architettura*, Electa, Milano;
- Norheim-Hagtun, I., Meier, P. (2010), *Crowdsourcing for Crisis Mapping in Haiti*, *Innovations: Technology, Governance*, 5(4): 81-89;
- Nuvolati, G. (2002), *Popolazioni in movimento città in trasformazione. Abitanti, pendolari, city users, uomini d'affari e flaneurs*, Il Mulino, Bologna;
- Nuvolati, G. (2007), *Mobilità quotidiana e complessità urbana*, Firenze University Press, Firenze;
- Nuvolati, G. (2011), (a cura di), *Lezioni di Sociologia urbana*, Il Mulino, Bologna;
- Nyerges, T.L., Couclelis, H. e McMaster, R. (2011), *The SAGE Handbook of GIS and Society*, Thousand Oaks, CA: SAGE;
- O'Brien, K., Hayward, B. e Berkes, F. (2009), *Rethinking social contracts: building resilience in a changing climate*. *Ecology and Society* 14(2): 12.
Disponibile online su <http://www.ecologyandsociety.org/vol14/iss2/art12/Synthesis>;
- O'Reilly, T. (2005), *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Disponibile on line su <http://www.oreillynet.com/lpt/a/6228>;
- Open Geospatial Consortium (2006), *OpenGIS Web Map Server Implementation Specification*;
- Open Geospatial Consortium (2007), *OpenGIS Geography Markup Language (GML) Encoding Standard*;
- Open Geospatial Consortium (2010), *OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*;
- Open Geospatial Consortium (2010), *OpenGIS Implementation Standard for Geographic information - Simple feature access - Part 2: SQL option*;
- Open Geospatial Consortium (2010), *OpenGIS Web Feature Service 2.0 Interface Standard*;

Open Knowledge Foundation 2012, Open Data Handbook Documentation. Release 1.0.0,
scaricabile dal sito: <http://opendatahandbook.org>;

Openshaw, S. (1984), *Ecological fallacies and the analysis of areal census data*. In 'Environment and Planning', A16, pp. 17-31;

Orban-Ferauge, F. (2011), *Participatory geographic information systems and land planning. Life experiences for people empowerment and community transformation*, Presses universitaires de Namur, Namur, Belgium;

Ortega, F., Gonzalez-Barahona, J.M., Robles, G., *The Top-Ten Wikipedias: A Quantative Analysis Using WikiXRay* (2007), ICSOFT '07, International Conference on Software and Data Technology, 46 – 53;

Ozemoy, V.M., Smith, D.R., Sicherman, A. (1981), *Evaluating computerized geographic information system using decision analysis*, Interfaces, 92-98;

Ozsu, M.T. (1994), *Query Processing Issues*, OO KBS;

Padgett, J.F. (1999), *Network Analysis and Social Structure*, Political Science 572, Sociology 585. Disponibile online sul sito: <http://home.uchicago.edu/~jpadgett/papers/syllabi/nt.pdf>;

Palen L., Anderson, K., Mark, G., Martin, J., Sicker, D., Palmer, M. e Grunwald, D. (2010), *A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters*, British Computer Society, 1-12;

Paolillo, L. (2011), *Documento di Piano del piano di governo del territorio ex art. 8 della Lr 12/2005 e smi*, Comune di Seveso, Politecnico di Milano;

Park, R. E., Burgess, E. W., McKenzie, R. D. (1925), *The city*, University of Chicago Press, Chicago;

Parker, R. N., Asencio, E. K. (2008), *GIS and spatial analysis for the social sciences: coding, mapping and modeling*. New York, Routledge;

Parmenter, B. (2007), *Spatial Analysis Tip Sheet – Spatial Joins and Overlays, Density, and Proximity Analysis*. Tufts University;

Parry, M. (2011), *Academics Join Relief Efforts Around the World as Crisis Mappers*, The Chronicle of Higher Education, disponibile sul sito: <http://chronicle.com/article/Academics-Join-Relief-Efforts/126912/#>;

- Pavlovskaya, M. E. (2002). *Mapping urban change and changing GIS: other views of economic restructuring*. *Gender, Place and Culture, A Journal of Feminist Geography* 9, pp. 281–289;
- Pavloskaya, M. E. (2004), *Other transitions: multiple economies of Moscow households in the 1990s*, *Annals of the Association of American Geographers* 94, pp. 329–351;
- Perencsik, A., Woo, S., Booth, B. (2004), *ArcGIS 9. Building a geodatabase*. Environmental Systems Research Institute, Redlands California;
- Petrov, A. (2012), *One Hundred Years of Dasyetric Mapping: Back to the Origin*, *Cartographic Journal*, Volume 49, Number 3, 256-264(9), Maney Publishing;
- Pfaff, R., Booth, B., Shaner, J. (2004), *ArcGIS 9. Editing in ArcMap*. Environmental Systems Research Institute, Redlands California;
- Pickles, J. (1995), Representation in an electronic age: geography, GIS, and democracy, in J. Pickles, *Groud Truth: The Social Implications of Geographic Information Systems*, New York, Guilford, pp. 1-30;
- Poese, I., Uhlig, S., Kaafar, M.A., Donnet, B. e Gueye, B. (2011), *IP Geolocation Databases: Unreliable?*, *ACM SIGCOMM Computer Communication Review*, 4(2): 53-56;
- Poincaré, H. (1908), *The foundation of science: Science and method*, tr. ingl. Lancaster, PA, The Science Press, 1946;
- Polgreen, P.M., Chen, Y., Pennock, D.M. e Nelson, F.D. (2008), *Using Internet Searches for Influenza Surveillance*, *Clinical Infectious Diseases*, 47 (11): 1443-1448;
- Pollock, J. (2011), *Streetbook: How Egyptian and Tunisian Youth Hacked the Arab Spring*, *Technology Review*, September/October 2011, Disponibile su <http://www.technologyreview.com/web/38379/>;
- Popper, K. R. (1963), *Congetture e confutazioni*, Ed. Il Mulino, Bologna;
- Preston, R.E. (1983), *La componente dinamica di Christaller Central Place Theory e il tema del cambiamento nella sua ricerca*, il geografo canadese, vol. 27, 4-16;

- Prigogine, I. (1996), *La fine delle certezze. Il tempo, il caos e le leggi della natura*, tr. it. Torino, Bollati Boringhieri;
- Prontera, F. (2003), *Tabula Peutingeriana, le antiche vie del mondo*, Leo S.Olschki Editore, Firenze;
- Rambaldi, G. (2006), *Resource use, development planning, and safeguarding intangible cultural heritage: lessons from Fiji Island*, in *Participatory Learning and Action*, 54, pp. 28-35;
- Residents of Hull House (1985), *Maps and Papers*, Thomas Y. Crowell & Co, Boston;
- Rigaux, P., Scholl, M., Voisard, A. (2002), *Spatial Databases with application to GIS*, Morgan Kaufmann;
- Ritterman, J., Osborne, M. e Klein, E. (2009), *Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic*, in Carrero, F.M., Gómez, J.M., Monsalve, B., Puertas, E. e Cortizo, J.C. (eds.), *1st International Workshop on Mining Social Media*, Sevilla, Spain pp. 9-18;
- Roebuck, K. (2013), *CAQDAS - Computer Assisted/Aided Qualitative Data Analysis: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits*, Emereo Pty Ltd, Volume 13, No. 2, Art. 12 – May 2012;
- Sakaki, T., Okazaki, M. e Matsuo, Y. (2010), *Earthquake shakes Twitter users: real-time event detection by social sensors*. Proceedings of the 19th international conference on World wide web: ACM, 851-860;
- Sauermann, L., Cyganiak, R. (2008), *Cool URIs for the Semantic Web*, W3C Interest Group Note. Disponibile su www.w3.org/TR/cooluris;
- Schiavi, A. (2002), *Vademecum cartografico*, V&P Università. Milano;
- Schlogel, K. (2009), *Leggere il tempo nello spazio. Saggi di storia e geopolitica*, Bruno Mondadori, Milano;
- Seidel, J. V. (1998), *Qualitative Data Analysis, Qualis Research*, testo disponibile online su <http://www.qualisresearch.com/Downloads/qda.pdf> ;

Shalizi C. R., A. C. Thomas. (2011), *Homophily and Contagion Are Generically Confounded*, in *Observational Social Network Studies, Sociological Methods & Research* May 2011, 40: 211-239;

Shi, X., Tseng, B. e Adamic, L.A. (2007), *Looking at the Blogosphere Topology through Different Lenses*, *Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007)*, Boulder, CO;

Shiffer, M. (2002), *Spatial Multimedia representation to support community participation*, in Craig, W., Harris, T. e Weiner, D., *Community Participation and Geographic Information Systems*, London, Taylor e Francis, pp. 309-320;

Silverman, D. (2002), *Come fare ricerca qualitativa*, Roma, Carocci;

Sørensen, P. (1998), *Theoretical Mechanism and the Empirical Study of Social processes*. In *Social Mechanism, An analytical Approach to Social Theory*, 238-266, Cambridge University Press. ;

Snow, J. (1854), *The cholera near golden-square and at Depford*. In 'Medical Times Gazette', 9, pp. 321-322;

Standby Task Force (2011), *The Security and Ethics of Live Mapping in Repressive Regimes and Hostile Environments*. Disponibile sul sito: <http://blog.standbytaskforce.com/?p=259>;

Starbird, K. e Palen, L. (2010), *Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Grammar to Support Citizen Reporting*. *Proceedings of the 7th International ISCRAM Conference (Short Paper)*, Seattle, WA;

Starbird, K. e Palen, L. (2011), *Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis*. *Conference on Computer Human Interaction (CHI 2011)*, Vancouver, BC, Canada: ACM;

Stefanidis, A., Crooks, A. e Radzikowski, J. (2011), *Harvesting Ambient Geospatial Information from Social media Feeds*, *Geojourna, Springer Science and Business Media*;

Stefanidis, A., Crooks, A. T., Radzikowski, J., Croitoru, A. e Rice M. (2011), *Social Media and Emergence of Open-Source Geospatial Intelligence*, in Tucker, C. and Tomes, R., *Socio-Cultural Dynamics and Global Security*:

Interdisciplinary Perspectives on Human Geography, an Era of Persistent Conflict, US Geospatial Intelligence Foundation (USGIF);

Steinberg, S.J. e Steinberg, S. (2006), *GIS for the Social Science*, Sage Publications, London;

Stengers, I. (2000), *The invention of Modern Science*, Smith D.W., Minneapolis, University of Minnesota Press;

Stevens Stanley Smith (1946), *On the theory of scales of measurement*, in *Science*, 03 (2684), p. 677–680;

Steward, D.W. (1984), *Secondary Research: Information, Sources And Methods*, Sage, London;

Stewart, J.Q. (1941), *An inverse distance variation for certain social influences*, *Science*, 93, 89-90;

Strati A. (1997), *La Grounded Theory*, in Ricolfi, L. (a cura di), *La ricerca qualitativa*, Roma, Carocci;

Strauss A. L., Corbin J. (1990), *Basic of Qualitative Research*. Grounded theory procedures and techniques, Sage, London;

Szegő, J. (1984), *A Census Atlas of Sweden*, Statistics Sweden, Central Board of Real Estate Data, Swedish Council of Building Research and The University of Lund, Stockholm;

Ticca, P.G. (2009), Sistemi informativi geografici e ricerca sociale, in Cannavò, L. e Frudà, L. (2009), *Ricerca sociale. Tecniche speciali di rilevazione, trattamento e analisi*, Carocci editore;

Tita, G. e Cohen, J. (2004), Measuring Spatial Diffusion of Shots Fired Activity Across City Neighborhoods. In Goodchild, M.F. e Janelle D.G., *Spatially Integrated Social Science.*, Oxford Press, New York;

Thomas, D. (2003), *UML - Unified or Universal Modeling Language? UML2, OCL, MOF, EDOC - The Emperor Has Too Many Clothes*, Journal of Object Technology, ETH Zurich, Chair of Software Engineering ©JOT, 2002 Vol. 2, N. 1, January-February 2003; disponibile online su www.jot.fm;

Tobler W. R. (1970), A computer movie simulating urban growth in the Detroit region. In *Economic Geography*, Vol. 46, Supplement: Proceedings. International

- Geographical Union. Commission on Quantitative Methods, pp. 234-240, published by Clark University;
- Tomaszewski, B. (2008), Producing Geo-historical Context from Implicit Sources: A Geovisual Analytics Approach. *Cartographic Journal* 45, 165-181;
- Tomaszewski, B. e MacEachren, A.M. (2010), Geo-Historical Context Support for Information Foraging and Sensemaking: Conceptual Model, Implementation, and Assessment. *IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2010)*, Salt Lake City, Utah, USA, 139-146;
- Travaglini, D. (2004), *Trasformazioni tra sistemi di coordinate: software disponibili, limiti e potenzialità*. *Forest@* 1(2), 128-134;
- Trobia, A. (2005), *La ricerca sociale quali-quantitativa*, Collana di Sociologia, FrancoAngeli, Milano;
- Trobia, A. (2001), *La sociologia come scienza rigorosa. Modelli simulativi, intelligenza collettiva, forme del mutamento*, FrancoAngeli, Milano;
- Trobia, A., Milia, V. (2011), *Social network analysis*, Carocci Editore, Roma;
- Ullman J. D. (1991), *Basi di dati e basi di conoscenza*, Jackson Libri, 1991;
- Ullmann , J.D & Widom, J. (1997), *Introduzione ai database*, Jackson Libri, 1997;
- Vieweg, S., Hughes, A., Starbird, K. e Palen, L. (2010), Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *Proceedings of the 28th international conference on Human factors in computing systems*: ACM, 1079-1088;
- Von Boventer, E. (1969), Luoghi di Walter Christaller centrali e aree periferiche: La Central Place Theory in Retrospect, *Journal of Regional Science* . Vol.9, 117-24;
- Voronoi, G. (1907), *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. *Journal für die Reine und Angewandte Mathematik*, 133:97-178, 1907;
- Wapner, S., Demick, J. (2002), The increasing contexts of context in the study of environment behavior relations, in Bechtel, R.B., Churchman, A. (Eds.), *Handbook of Environmental Psychology*, pp. 3-14;

- Warf, B., Arias, S. (2008), *The Spatial Turn. Interdisciplinary Perspectives*, New York: Routledge, London;
- Wasserman, S., Faust, K. (1994), *Social network analysis: methods and application*, Cambridge University Press, Cambridge;
- Weber, M. (1961), *La città*, in Id, *Economia e Società*, 2 voll., Milano, Edizioni di Comunità, vol. II;
- Weiner D. e Harris T. (2003), *Community-integrated GIS for land reform in South Africa*, *The URISA Journal*, 15, pp. 61-73;
- Wilson, A.G. (1970), *Entropy in Urban and Regional Modeling*, Pion, London;
- Wirth, L. (1945), *Human ecology*, in *The American Journal of Sociology*, L, 438-488;
- Woodruff , A. e Wallace, T. (2012), *Bostonography*, <http://www.bostonography.com> ;
- Worboys, M.F. (1995), *GIS, a computing perspective*, Taylor & Francis;
- Wright, J.K. (1938), Problems in population mapping, in *Notes on Statistical Mapping, With Reference to the Mapping of Population Phenomena*, AGS and Population Association of America, New York;
- Yingwei Cui (2001), *Lineage Tracing in Data Warehouses*, PhD thesis, Stanford University;
- Yourdon E. – *Analisi strutturata dei sistemi. Concetti e metodi.* – Jackson, 1990;
- Zadeh, L.A. (1965). *Fuzzy sets and systems*. In ‘System Theory’, Fox J. editor, Brooklyn, NY, Polytechnic Press, pp. 29-39;
- Zajczyk, F. (1996), *Fonti per le statistiche sociali*. Franco Angeli, Milano;
- Zerubabel, E. (1992), *Terra Cognita. The Mental Discovery of America*, Rutgers University Press, New Brunswick;
- Zhang, J., Ackerman, M.S. e Adamic, L. (2007), *Expertise Networks in Online Communities: Structure and Algorithms*, Proceedings of the 16th International Conference on World Wide Web, Banff, Canada, pp. 221-230;
- Zipf, G.K. (1949), *Human Behavior and the principle of Least Effort*, Hafner, New York;

- Zlatev, J. (2007), Spatial Semantics, in Cuyckens, H. e Geeraets, D., *Handbook of Cognitive Linguistics*, Oxford University Press;
- Zook, M., Graham, M., Shelton, T. e Gorman, S. (2010), *Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake*, World Medical & Health Policy 2, Article 2;
- Zourbagh, H.W. (1995), Le aree naturali della città, in R. Rauty (a cura di), *Società e metropoli. La scuola sociologica di Chicago*, Roma, Donzelli, 97-102.