

Open Data: una sperimentazione di dati spaziali ad accesso libero

Roberto della Maggiore, Daniela Nuvolone, Stefano Cresci, Roberto Fresco

Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI) CNR Via Moruzzi 1, I56124, Pisa
Tel. 050 3152944, Fax 050 3152040, e-mail Roberto.dellaMaggiore@isti.cnr.it

Riassunto

La mancata disponibilità di dati cartografici costituisce spesso un ostacolo alla realizzazione di progetti su tematiche GIS nei quali è prevista l'analisi di dettaglio di fenomeni spaziali. Un argomento particolarmente dolente, nella realtà italiana, è la georeferenziazione degli indirizzi civici; infatti, salvo eccezioni ancora piuttosto rare, le pubbliche amministrazioni non dispongono di mappature complete e affidabili di toponomastica e numerazione civica. Per contro, alcune imprese commerciali che operano nel settore dell'informazione geografica possiedono dati di tal genere e rendono disponibili gratuitamente le coordinate corrispondenti agli indirizzi civici, con possibilità di importarle in una applicazione utente.

L'obiettivo del presente lavoro è la sperimentazione di uso di dati liberamente scaricati dalla rete; il caso di studio è un progetto sugli effetti del traffico sulla popolazione scolastica di Palermo: gli studenti partecipanti, di cui si conosce l'indirizzo di abitazione, devono essere mappati sul territorio. La geocodifica degli indirizzi è stata ottenuta accedendo via *web* ad un servizio commerciale (*Google Maps*), tramite l'API (*Application Programming Interface*), fornita dal servizio stesso. Per l'accesso è stato predisposto un apposito codice *Javascript*. Dei 715 indirizzi richiesti si è ottenuta risposta per 700, 673 dei quali (circa il 94% del totale) sono stati ritenuti affidabili.

Abstract

Many GIS-based projects are often hindered by the lack of cartographic data, mainly in case of high-detailed analysis of spatial phenomena. A very sore point of Italian scenario is address georeferencing; in fact, apart from a few number of local cases, Italian public administrations do not have yet produced complete and affordable numerical cartographic data. On the other hand some commercial companies, which operate in the field of the Geographic Information, own such data and they allow free download of geographic coordinates of civic addresses, with the possibility of importing them in a user application.

This study aims to test the use of free data, downloaded from the Internet; the case study is a project on the effects of traffic-related air pollution on respiratory health of a schoolchildren sample living in Palermo (Sicily). The students participating in the survey provided their residence addresses. Addresses geocoding was made gaining access via web to a commercial service (Google Maps), by means of an API (Application Programming Interface), provided by the service itself. For the access a Javascript code was implemented. The service provided the coordinates for 700 addresses out of the 715 requested; we accepted 673 addresses (94% of the request).

Introduzione

La cartografia di base è un dato di fondamentale importanza per chi si pone l'obiettivo di analizzare da un punto di vista spaziale gli eventi e i fenomeni che riguardano un territorio. La possibilità di localizzare spazialmente con precisione un dato raccolto sul campo e di riferirlo alla realtà circostante può costituire una discriminante nella decisione di effettuare o meno uno studio. Anche per questi motivi stiamo assistendo negli ultimi tempi anche in Italia e in Europa ad un notevole fermento nell'ambiente legato all'informazione geografica che spinge per la libera fruizione dei dati

oltre che del software per gestirli (GFOSS, ForumIGT, The Open Knowledge Foundation). Negli Stati Uniti è attiva un'organizzazione (Open Data Consortium) che definisce le modalità di accesso ai dati geografici e la gestione dei diritti di proprietà sui dati stessi (Joffe, 2003). D'altra parte, accanto a prestigiosi enti istituzionali, anche le maggiori imprese di produzione a livello mondiale di software GIS e di sviluppo di dati spaziali aderiscono all'Open Geospatial Consortium (OGC) che pone l'interoperabilità (ma non la libera condivisione) dei dati spaziali fra gli obiettivi principali dell'organizzazione.

Per quanto riguarda i dati cartografici numerici di dettaglio, in Italia sono tipicamente gli enti locali che ne detengono il patrimonio, anche se la situazione non è affatto omogenea: accanto a regioni e comuni che possono vantare situazioni ben consolidate e radicate nel tempo, altri enti della Pubblica Amministrazione stanno muovendo i primi passi verso una informatizzazione del dato geografico. Parallelamente molte iniziative private, legate a realtà estese a livello mondiale, dispongono di propri dati, indubbiamente di buona qualità, ovviamente vincolati da interessi commerciali per quanto riguarda le applicazioni più rilevanti. Queste imprese tuttavia offrono anche un numero incredibile di servizi gratuiti, fruibili via Internet, in alcuni casi mettendo anche a disposizione il dato in forma elaborabile dall'utente.

La direttiva Inspire

Recentemente approvata dalla Comunità Europea (14 marzo 2007), la direttiva Inspire (*IN*frastructure for *S*patial *I*nfoRmation in *E*urope) (Gazzetta, 2007) riporta fra le considerazioni iniziali l'auspicio che i dati territoriali prodotti dalla amministrazione pubblica si possano "ricercare, convertire, consultare e scaricare" mediante una rete di servizi atta a garantire l'interoperabilità dei dati. E infatti le disposizioni che la direttiva emana vanno nella direzione della condivisione dei dati geografici a livello europeo, stabilendo preliminarmente che per dati e servizi vengano costituiti metadati omogenei che ne rendano agevole l'accesso e chiare le condizioni di utilizzo. La direttiva detta inoltre le condizioni di interoperabilità di dati e servizi, fissando quali servizi debbano essere implementati per i dati pubblicati e stabilisce che almeno la ricerca e la consultazione dei dati siano rese disponibili gratuitamente al pubblico. Come si vede da questo brevissimo accenno, Inspire regola le modalità di produzione, di pubblicazione e di distribuzione dei dati, mentre nessun obbligo viene imposto agli stati membri riguardo alla produzione dei dati stessi: è perciò ragionevole aspettarsi che l'entrata in vigore della disposizione rallenti ulteriormente il processo di costituzione dei dati da parte delle pubbliche amministrazioni che sono ora tenute ad una esecuzione molto completa e capillare, oltre a dover provvedere all'adeguamento alla nuova direttiva dei dati esistenti. Pertanto la sperimentazione di uso di dati proprietari (ma di libero accesso) che viene presentata in questo articolo pare degna di nota e indica una possibilità pratica destinata a valere per qualche tempo ancora (la direttiva Inspire esprime anche la consapevolezza che alcuni set di dati territoriali e di servizi ad essi relativi "sono detenuti e gestiti da terzi" e stabilisce le modalità con cui questi potranno essere inseriti nel circuito Inspire, previa omologazione col dato pubblico).

Il caso di studio

Il lavoro presentato fa parte di un progetto il cui scopo è quello di analizzare lo stato di salute respiratoria di un campione della popolazione scolastica della città di Palermo. Una campagna di raccolta di dati epidemiologici è stata svolta allo scopo in alcune scuole cittadine. L'analisi statistico-epidemiologica deve essere effettuata facendo riferimento alla quantità di traffico veicolare presente in prossimità delle scuole e delle abitazioni degli scolari; ciò allo scopo di evidenziare eventuali effetti sulla salute delle emissioni nocive prodotte dalla combustione dei carburanti. I dati cartografici per l'implementazione del progetto sono quindi costituiti dal reticolo stradale e dalla ubicazione delle scuole e delle abitazioni degli studenti. Per il primo aspetto non sussistono problemi, essendo disponibile presso il comune di Palermo la cartografia numerica relativa alle strade; non esiste invece la georeferenziazione di numeri civici. Si è fatto perciò ricorso

al servizio di geocodifica di indirizzi *Google Maps* offerto gratuitamente dalla società di fornitura di servizi Internet *Google Inc.* Il modo più comune e più noto di utilizzare tale servizio consiste nell'inviare, tramite l'interfaccia on-line appositamente predisposta da *Google Maps*, l'indirizzo di cui si desidera conoscere l'ubicazione: il sistema risponde visualizzando il punto corrispondente su di una mappa stilizzata; con questo metodo le coordinate dell'indirizzo non vengono fornite in chiaro al richiedente. In alternativa è possibile accedere al servizio di geocodifica tramite un servizio accessorio di *Google Maps* che fornisce un'API (*Application Programming Interface*), pubblicata su *web*, con la quale si ottengono i valori numerici delle coordinate. Per essere ammessi al *download* occorre registrarsi presso Google ed ottenere una "*Maps API key*" di autorizzazione. La registrazione è gratuita; l'unica limitazione imposta è sul numero di richieste di geocodifica, fissato in 50.000 al giorno, cioè al massimo una richiesta mediamente ogni 1,73 secondi.

Occorre precisare subito che se da un lato *Google Maps* fornisce l'accesso all'API senza nessun vincolo, dall'altro non pubblica metadati che possano servire all'utilizzatore per impostare con consapevolezza il suo lavoro (tutto ciò che siamo riusciti a sapere è che "*Google Maps* si basa su standard internazionali ampiamente riconosciuti per le convenzioni di denominazione e mappatura", segue il riferimento alla norma ISO-3166). Nel caso delle coordinate corrispondenti ad un indirizzo richiesto il servizio restituisce un codice di accuratezza che può essere utilizzato per fare una prima valutazione sul dato ottenuto, ma l'esperienza insegna che in generale l'approccio migliore consiste nel verificare in mappa la congruenza del dato. Per l'Italia i dati di geocodifica forniti da *Google Maps* sono per la maggior parte prodotti da *TeleAtlas* che li concede in licenza a *Google*. *TeleAtlas* assicura una copertura dell'84,48 per cento del totale degli indirizzi del territorio italiano. Il metodo seguito da *TeleAtlas* per la georeferenziazione dei numeri civici consiste in genere nell'interpolare i civici stessi lungo il tratto di strada a cui un certo *range* di numeri appartiene, quindi più corto è il tratto, maggiore è la precisione di geocodifica.

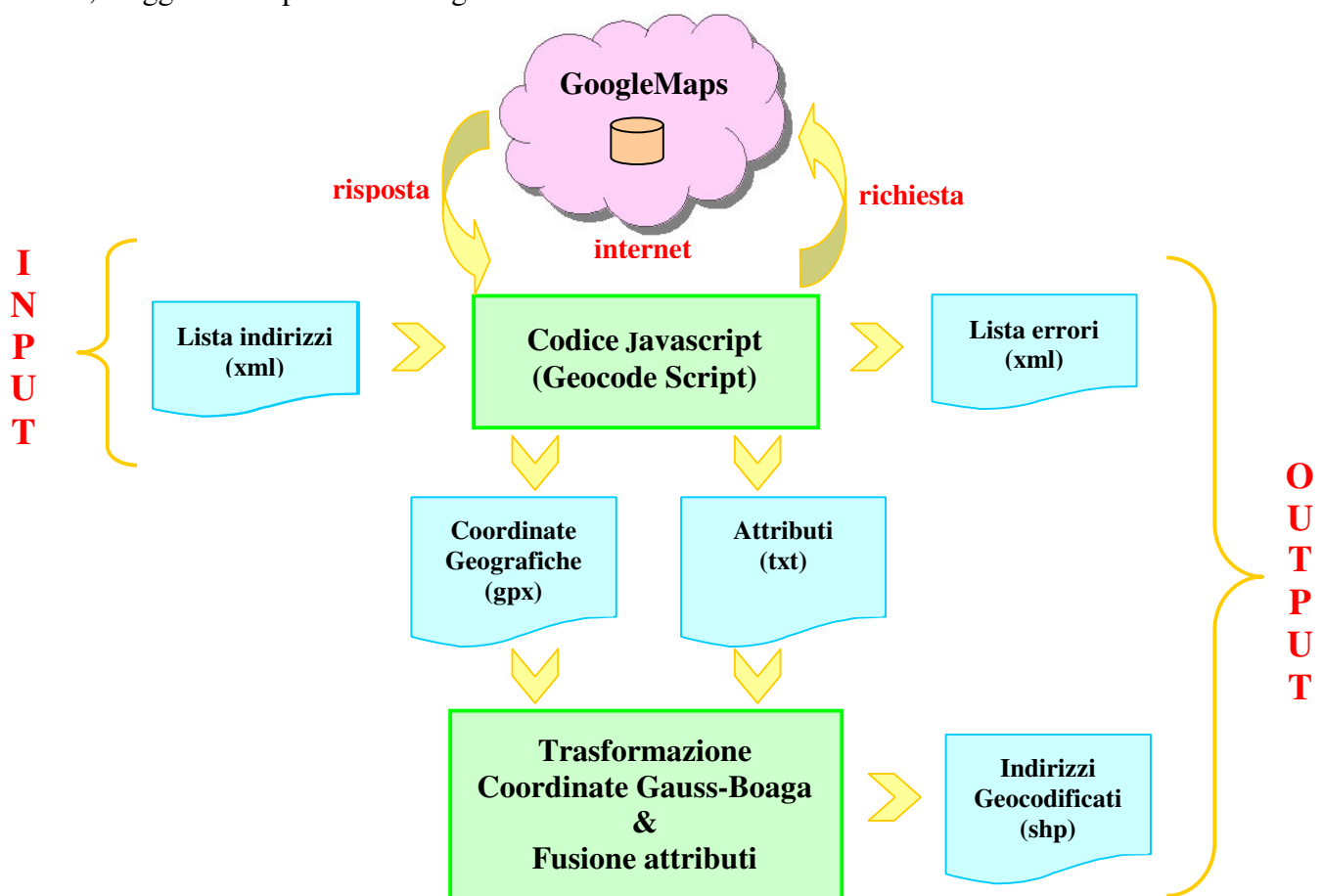


Figura 1 – Schema del processo di geocodifica degli indirizzi

Per la realizzazione del progetto è stato sviluppato il codice *Geocode*, uno *script* che accede a *Google Maps* tramite l'API per ottenere le coppie di coordinate corrispondenti agli indirizzi contenuti in una lista.

La figura 1 sintetizza il processo di trasformazione descritto di seguito. La lista degli indirizzi è stata inserita in un file di tipo XML (figura 2) che è sottoposto a *Geocode* per l'elaborazione. Il programma apre la sessione con *Google maps* presentando le credenziali di autorizzazione, quindi invia le richieste di geocodifica una per volta, sincronizzandosi con il servizio. *Geocode* produce tre distinti file in uscita: uno contiene le coordinate degli indirizzi geocodificati (file di tipo gpx, figura 3), il secondo i relativi attributi (indirizzo in chiaro e codice di accuratezza – da 1 a 8 -, file di tipo txt, figura 4), mentre nel terzo vengono registrate le segnalazioni di errore per gli indirizzi che non vengono riconosciuti dal servizio (figura 5). Per ogni risposta corretta che riceve, *Geocode* crea un nuovo record sia nel file delle coordinate che in quello degli attributi, assegnando ad essi lo stesso identificativo.

```
<marker>
<marker addr="114 via Catalano Agostino, Palermo, it" id="1745"/>
<marker addr="34 via Drago Aurelio, Palermo, it" id="549"/>
<marker addr="9 via Holm Adolfo, Palermo, it" id="1676"/>
<marker addr="37 via Millo Ammiraglio, Palermo, it" id="1683"/>
<marker addr="45 via Millo Ammiraglio, Palermo, it" id="1671"/>
<marker addr="11 via Raccuglia Antonio, Palermo, it" id="1044"/>
<marker addr="8 via Rosmini Serbati Antonio, Palermo, it" id="1800"/>
```

Figura 2 – Stralcio di lista di indirizzi da geocodificare

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<gpx version="1.1" creator="Geocode Script Version 4.3">
  <wpt lat="38.103552" lon="13.340827"><name>1745</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.105251" lon="13.33154"><name>549</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.127804" lon="13.330785"><name>1676</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.122111" lon="13.336867"><name>1683</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.122294" lon="13.336786"><name>1671</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.077775" lon="13.404924"><name>1044</name><sym>Waypoint</sym></wpt>
  <wpt lat="38.191905" lon="13.285258"><name>1800</name><sym>Waypoint</sym></wpt>
```

Figura 3 – Stralcio di file con le coordinate restituite da Google Maps

```
name,ind_corto,ind_lungo,accuracy
1745, 114 via Catalano Agostino, "Via Agostino Catalano, 90129 Palermo, Palermo (Sicilia), Italy", 6
549, 34 via Drago Aurelio, "Via Aurelio Drago, 34, 90129 Palermo, Palermo (Sicilia), Italy", 8
1676, 9 via Holm Adolfo, "Via Adolfo Holm, 9f, 90145 Palermo, Palermo (Sicilia), Italy", 8
1683, 37 via Millo Ammiraglio, "Via Ammiraglio Millo, 37, 90145 Palermo, Palermo (Sicilia), Italy", 8
1671, 45 via Millo Ammiraglio, "Via Ammiraglio Millo, 45, 90145 Palermo, Palermo (Sicilia), Italy", 8
1044, 11 via Raccuglia Antonio, "Via Antonio Raccuglia, 11, 90124 Ciaculli, Palermo, Palermo (Sicilia), Italy", 8
1800, 8 via Rosmini Serbati Antonio, "Via Antonio Rosmini Serbati, Palermo, Palermo (Sicilia), Italy", 6
```

Figura 4 – Stralcio di file con le informazioni attributali restituite da Google Maps

```
<markers>
<marker addr="57 via A. Sacha, Palermo, it" err_reason="Unkown Address" err_code="602"/>
<marker addr="11 via Callia, Palermo, it" err_reason="Unkown Address" err_code="602"/>
<marker addr="0 Contrada Scozzari, Palermo, it" err_reason="Unkown Address" err_code="602"/>
<marker addr="89 via Cuxino, Palermo, it" err_reason="Unkown Address" err_code="602"/>
<marker addr="9 fondo Napoletano, Palermo, it" err_reason="Unkown Address" err_code="602"/>
<marker addr="17 via Marinai D'Italia, Palermo, it" err_reason="Unkown Address" err_code="602"/>
```

Figura 5 – File con le indicazioni di errore restituite da Google Maps

Google Maps restituisce coordinate geografiche (WGS 84) nel formato gradi e centomillesimi di grado. Quando tutti gli indirizzi della lista sono stati elaborati *Geocode* termina fornendo alcuni dati statistici sul lavoro svolto.

Le coordinate così ottenute devono essere convertite nel sistema Gauss Boaga per renderle compatibili con i dati cartografici di Palermo. La conversione è stata effettuata utilizzando il software di uso gratuito *GPS Utility*, con il quale è possibile ottenere un file in formato shape. Infine il processo di geocodifica degli indirizzi è stato completato ricorrendo ad un software GIS per operare la fusione fra il file contenente la sola informazione geografica prodotto da *GPS Utility* ed il file degli attributi prodotto da *Geocode* (per la sperimentazione è stato utilizzato ArcView di ESRI).

Come detto, il procedimento sopra descritto è stato applicato alla geocodifica di 715 indirizzi. Gli indirizzi ottenuti in risposta da *Google Maps* sono stati analizzati singolarmente in mappa tenendo conto anche del codice di accuratezza. La tecnica utilizzata è stata di tipo *trial and error*, infatti alcuni indirizzi sono stati riformulati dopo le prime risposte, di altri si è deciso l'eliminazione perché la risposta pareva del tutto inverosimile. Il numero degli indirizzi geocodificati ritenuti affidabili è 673. Il risultato complessivo del lavoro svolto è sintetizzato nella mappa di figura 6; la figura 7 riporta invece un dettaglio dal quale si può visivamente apprezzare la qualità dei dati.

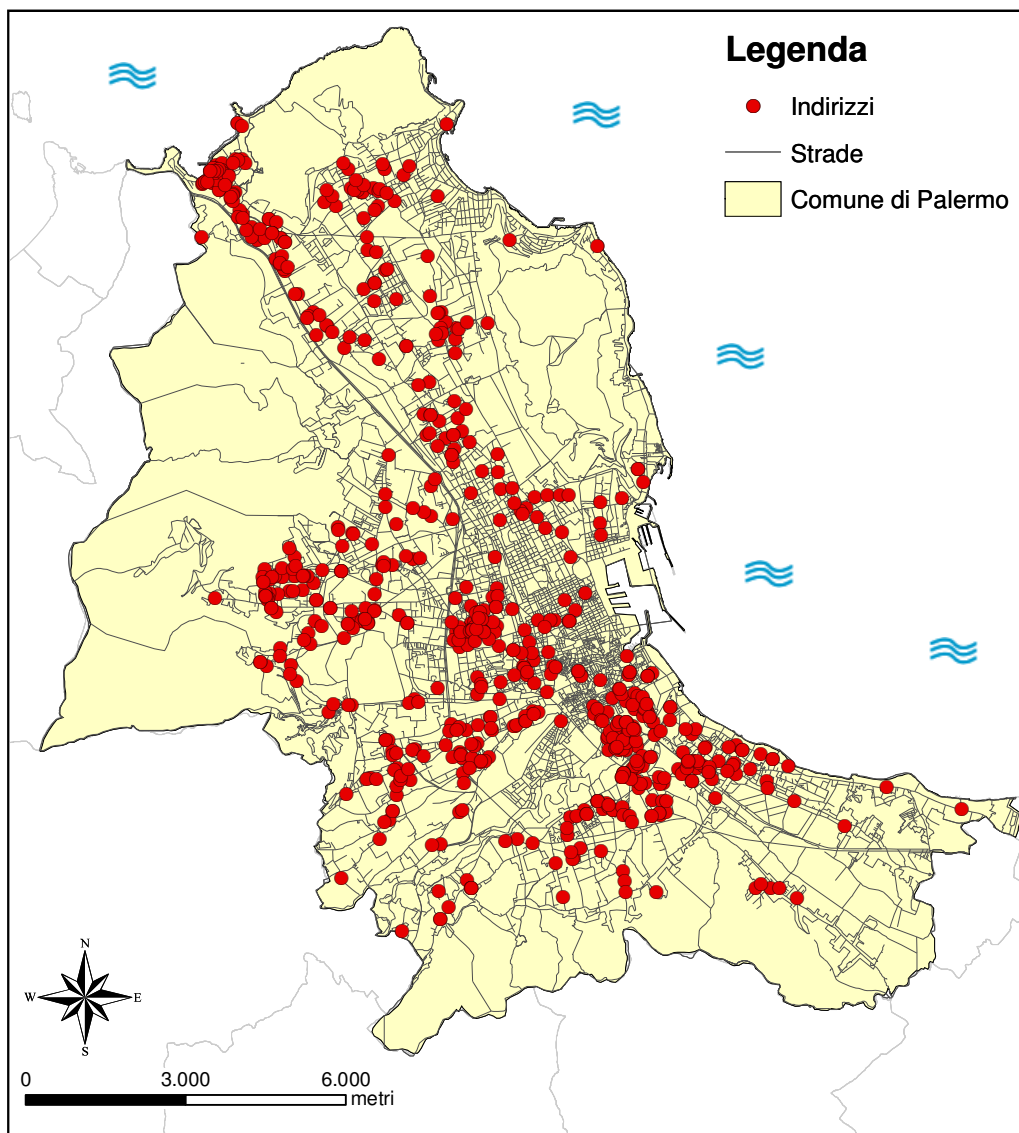


Figura 6 – Mappa degli indirizzi geocodificati

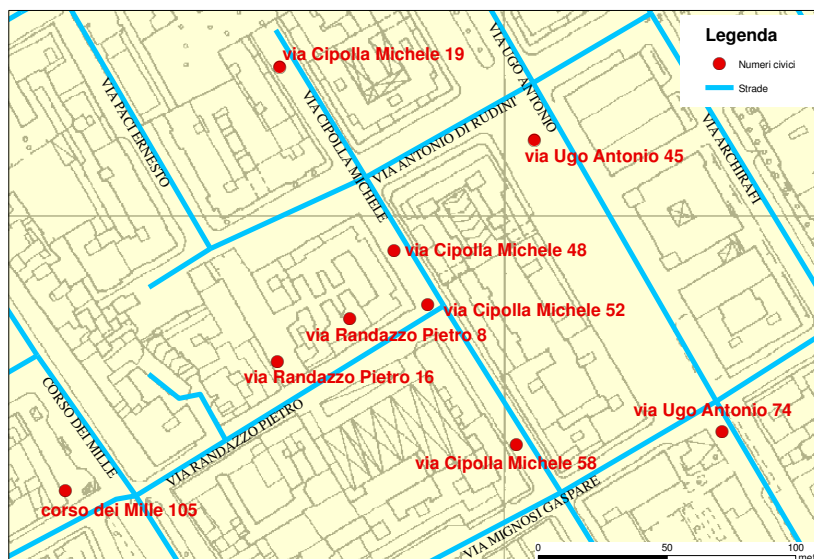


Figura 7 – Dettaglio della geocodifica

Conclusioni

L'occasione che ci ha portato a mettere in atto la sperimentazione presentata in questo articolo è stata la necessità di georeferenziare una quantità notevole, anche se non elevatissima, di indirizzi in una città italiana per la quale ancora la pubblica amministrazione non dispone della geocodifica dei numeri civici. La sfida a cui far fronte era duplice: il progetto necessitava della mappatura degli indirizzi e questo doveva avvenire in tempi ragionevolmente brevi. Il ricorso a *Google Maps* ha consentito di raggiungere entrambi gli obiettivi, con in più il piacevole particolare di non aver dovuto investire fondi nell'operazione.

L'esito positivo della sperimentazione ha anche una portata di carattere generale. Infatti ha mostrato come, in un settore in rapida evoluzione come quello della libera circolazione dei dati, sia possibile ottenere, liberamente e gratuitamente, dati geografici in formato elaborabile, anche quando questi ancora non sono in produzione presso la pubblica amministrazione. E ciò a fronte delle azioni intraprese dalle istituzioni che, per necessità imposta da requisiti di completezza oltre a fattori innati nel settore, procedono con la consueta lentezza.

Bibliografia

ForumIGT - <http://www.forumigt.it>

Gazzetta ufficiale dell'Unione europea, GU L 108 del 25.4.2007

GFOSS - <http://www.gfoss.it>

Joffe Bruce A. (2003), Model Data Distribution Policy, Open Data Consortium project (http://www.opendataconsortium.org/documents/Data_Policy-4b.pdf)

OGC - Open Geospatial Consortium - Geospatial Digital Rights Management Reference Model (GeoDRM RM), 2006, http://portal.opengeospatial.org/files/?artifact_id=14085

The Open Knowledge Foundation - <http://www.okfn.org>