

Condividere il patrimonio cartografico tramite un sistema basato su un core di metadati geografici

Vincenzo DEL FATTO, Luca PAOLINO, Monica SEBILLO, Genny TORTORA, Giuliana VITIELLO

Dipartimento di Matematica e Informatica
Università di Salerno
Via Ponte don Melillo
84084 Fisciano (SA) - Italy
{vdelfatto, lpaolino, msebillo, tortora, gvitiello}@unisa.it

RIASSUNTO

Un aspetto cruciale nello sviluppo di progetti in ambito GIS è l'alto costo che occorre affrontare per raccogliere i dati geografici. D'altra parte, esiste una grande mole di dati immediatamente disponibili in Internet che possono essere visualizzati o direttamente scaricati. In questo lavoro presentiamo una metodologia per la ricerca e la condivisione dei dati, basata sulla gestione di un insieme di metadati geografici. Tale metodologia è stata sperimentata in un progetto pilota, le cui caratteristiche principali sono descritte in questo lavoro.

ABSTRACT

A crucial aspect in GIS project development is the high cost that companies face to collect geographic data. On the other hand, GIS data are abundant and readily available on the Internet for either on-line viewing or direct downloading. In this paper we introduce a method for retrieving and sharing data, based on a minimal set of geographic metadata. Moreover, the system developed in order to validate the method is described along with its basic features.

KEYWORDS: *Geographical Information System, Data Sharing, Metadata, P2P*

LA CONDIVISIONE DEI DATI

Quando si utilizza un motore di ricerca on-line per trovare una cartografia, le tradizionali tecniche di interrogazione non sono più sufficienti, poiché l'uso dei criteri di ricerca basati su parametri come data o nome del file consente agli utenti di formulare delle query che discriminano i dati solo sulla base delle loro proprietà organizzative. Diversamente, la duplice natura del dato geografico, descrittiva e spaziale, determina un'implicita complessità nell'attività di gestione (Longley et al., 2001). Infatti, gli utenti esperti di contenuti cercano i dati di interesse tenendo conto del contenuto dei file e della loro descrizione, per cui un risultato è soddisfacente quando fornisce conoscenza geografica in tal senso.

Sulla base di queste considerazioni è allora importante definire un metodo che tenga conto delle proprietà riguardanti il contenuto e la descrizione dei dati, così come dei modelli e delle strutture sottostanti usati per la loro rappresentazione, per poter dare agli utenti la capacità di catturare tutte le peculiarità del dato geografico.

A questo scopo, abbiamo approfondito lo studio riguardante gli standard relativi ai metadati, in particolare, ISO/FDIS 19115, Dublin Core e Repertorio Nazionale CNIPA, con l'obiettivo di individuare ed estrarre un insieme minimale di elementi significativi e adatti alla descrizione del contenuto delle risorse geografiche.

Lo studio è stato condotto coinvolgendo una equipe eterogenea di esperti¹ che lavorano in domini in cui le applicazioni GIS sono determinanti sia per scopi di visualizzazione che di analisi. Abbiamo intervistato botanici, cartografi, veterinari, geologi ed archeologi, ed abbiamo chiesto loro di indicare e pesare i criteri essenziali sulla base dei quali effettuano la ricerca di risorse. Le risposte al questionario ha messo in risalto due aspetti fondamentali:

- alcuni elementi giocano un ruolo determinante durante i processi di ricerca,
- quando è effettuata da un esperto, la valutazione della qualità di un documento dipende anche dalle sue proprietà strutturali, come il formato del file, il sistema di riferimento, la proiezione, ecc..

Il risultato ottenuto in questa prima fase è quindi l'individuazione di 17 attributi corrispondenti ad una *core* di metadati, che rispondono alle seguenti domande: "Esiste un dataset su uno specifico argomento (*cosa*)?", "per uno specifico luogo (*dove*)?", "per una specifica data o periodo (*quando*)?", "dove è la corrispondente risorsa?". La Tabella 1 riporta i metadati selezionati, la loro descrizione e i corrispondenti domini espressi in accordo agli standard considerati in partenza.

Inoltre, è utile notare che gli esperti coinvolti in questo studio, in maniera concorde hanno riconosciuto che tra gli elementi individuati, è possibile estrarre un sottoinsieme minimale necessario a caratterizzare tutte le risorse indipendentemente dallo specifico dominio a cui ci si riferisce. Nella tabella questi elementi sono indicati con un asterisco.

Attributo	Descrizione	Dominio
Titolo *	Nome attraverso il quale la risorsa è conosciuta	Testo libero
Keywords*	Parole chiave rappresentanti il contenuto della risorsa	Testo libero
Abstract	Breve riassunto del contenuto della risorsa	Testo libero
Autore*	Un'entità responsabile del contenuto della risorsa	Testo libero
Lingua	La lingua usato nel dataset	Valori predefiniti (ISO 639-2)
Tipo di rappresentazione	Metodo usato per rappresentare il contenuto della risorsa (Vector, Raster, ...)	Valori predefiniti
Formato*	Formato del dato da distribuire (SHP, DWG, TIF, ...)	Valori predefiniti
Geometria	Un valore che indica la rappresentazione geometrica (punto, linea, regione, ...)	Valori predefiniti
Tipo di sistema di riferimento	Tipo di sistema di riferimento (Geografico, proiettato, ...)	Valori predefiniti

¹ Professionalità partecipanti al Centro Regionale di Competenza per lo Sviluppo ed il Trasferimento dell'Innovazione applicata ai Beni Culturali e Ambientali - INNOVA, progetto triennale finanziato dalla Regione Campania.

Sistema di riferimento*	Nome del sistema di riferimento	Testo libero
Copyright	Informazioni riguardanti il copyright associato alla risorsa	Testo libero
Data*	Data di riferimento per la risorsa	AAAA/MM/GG (ISO 8601)
Scala *	Fattore che fornisce una generale comprensione della densità dei dati spaziali nel dataset.	Testo libero
Locazione della risorsa	Posizione fisica dove la risorsa può essere trovata.	Testo libero
Locazione geografica del dataset	Informazioni riguardanti il bounding box del dataset e la relative estensione temporale.	Testo libero
Responsabile dei dati	Individuazione delle persone e delle organizzazioni associate alla risorsa	Testo libero
Risorsa on-line	Informazioni riguardo la sorgente online dalla quale la risorsa può essere ottenuta.	Testo libero

Tabella 1: il core di metadati

La Tabella 2 riporta la corrispondenza tra gli elementi individuati e gli standard studiati, mettendo in risalto il modo in cui la realtà italiana analizzata richieda un adattamento della proposta internazionale, mentre sia molto vicina al core del CNIPA. Le diversità derivano dagli specifici requisiti del dominio coinvolto.

Anche in questo caso l'asterisco indica l'appartenenza dell'elemento al core dello standard corrispondente, mentre *Ob*, *Op* e *Cond* indicano rispettivamente che l'elemento è obbligatorio, opzionale o condizionato.

Attributo	Dublin Core	ISO 19115	Repertorio Nazionale CNIPA
Titolo *	Title	Dataset title (Ob)*	Titolo (Ob) *
Keywords*	Subject	Keyword (Ob)	Parole chiave (Ob) *
Abstract	Description	Abstract describing the dataset (Ob) *	Descrizione (Ob) **
Autore*	Creator	Dataset responsible party (Op) * (when role="originator")	Nome dell'Ente (Ob) * (con ruolo = "autore")
Lingua	language	Dataset language (Ob) *	Lingua dei dati (Ob) *
Tipo di rappresentazione	Type	Spatial representation type (Op) *	Tipo di rappresentazione spaziale (Con) *
Formato*	Format	Distribution format (Op) *	Formato (Ob) *
Geometria	N/A	N/A	N/A

Tipo di sistema di riferimento	N/A	N/A	N/A
Sistema di riferimento*	N/A	Reference System (Op) *	Sistema di riferimento spaziale (Ob) *
Copyright	N/A	N/A	N/A
Data*	Date	Dataset reference date (Ob) *	Data (Ob) *
Scala *	N/A	Equivalent Scale (Con)	Scala equivalente (Con) *
Locazione della risorsa	Source	Statement (Con)	Genealogia del dato – Processo di Produzione (Ob) *
Locazione geografica del dataset	Coverage: Spatial	Geographic location of the dataset (by four coordinates or by geographic identifier) (Con) *	Localizzazione Geografica dei dati (Ob) *
Responsabile dei dati	Publisher	Dataset responsible party (Op) * (when role=" publisher ")	Nome dell'Ente (Ob) (con ruolo = "editore") *
Risorsa on-line	Identifier	Linkage (Ob) *	Sito (O)

Tabella 2: corrispondenza tra i metadati

Un primo utilizzo di questa classificazione ha riguardato un sistema di condivisione del patrimonio cartografico nell'ambito del progetto INNOVA. Utenti autorizzati all'accesso su una Intranet possono interagire con un database cartografico, su cui sono consentite sia operazioni di download che di upload. In particolare, nel caso di operazioni di ricerca finalizzate al download di risorse, l'utente può selezionare diversi criteri di ricerca, tutti relativi ai metadati ed al contenuto del file di interesse. Ciò garantisce la bontà del file restituito sia in termini di contenuto che di struttura e organizzazione dei dati. Viceversa, nel caso di una operazione di upload, viene chiesto all'utente di compilare la scheda relativa al core di metadati, così da avere sempre una descrizione completa della risorsa e garantire le successive operazioni di ricerca e di download.

Il sistema proposto è stato a lungo testato ed attualmente rappresenta uno strumento di supporto alle attività di scambio svolte dai ricercatori di INNOVA che partecipano ad iniziative nell'ambito dei sistemi informativi geografici. In particolare, il periodo di validazione ha evidenziato la necessità di opportune integrazioni, con l'obiettivo rendere il sistema più performante per le funzionalità messe a disposizione dell'utente e più immediato da un punto di vista dell'usabilità.

Innanzitutto, a partire dalla classificazione introdotta è possibile apportare ulteriori raffinamenti al sistema di ricerca, sia in termini di tipi di query che di valutazione del risultato ottenuto. Di fatto, le query convenzionali consentono solo di verificare la corrispondenza tra pattern, e i risultati ottenuti sono tipicamente organizzati sulla base della percentuale di corrispondenza verificata. Diversamente, l'esperienza maturata dagli esperti ha messo in risalto la necessità di associare un peso ad ogni metadato sulla base dell'importanza che questo riveste nell'ambito della ricerca. Tipicamente, i pesi sono associati tenendo conto di una tipologia generica di utenti che

possono avere degli interessi generici. Tuttavia, talvolta è necessario considerare dei pesi dipendenti dal dominio di applicazione, soprattutto quando gli utenti interessati rappresentano una comunità di esperti. In questo caso alcuni metadati discriminano fortemente i dati geografici e la conoscenza derivata da essi differisce notevolmente nonostante il contenuto della risorsa. Basti pensare ad esempio alla diversa rilevanza che i botanici e i geologi assegnano al concetto di data.

Scenari che riguardano attività botaniche evolvono molto rapidamente, a differenza di quelli riguardanti gli aspetti geologici. Di conseguenza un processo di ricerca richiesto da un geologo può essere eseguito tenendo conto solo marginalmente dell'elemento data.

Un ulteriore miglioramento concerne i tipi di query che un utente può formulare. A tale scopo, abbiamo definito una metodologia di ricerca che consente di interrogare i dati con due tipi di selezione diversi ma tra loro integrabili, *Basic / Advanced matching* e *Superficial / Deep level*. In particolare, per quanto riguarda la ricerca, essa può essere effettuata per livelli, vale a dire, il file che contiene i metadati può essere analizzato solo per ciò che riguarda i metadati, oppure anche per quanto riguarda i dati e quindi il contenuto della risorsa ad esso associata. Per quanto riguarda invece il tipo di matching richiesto, la metodologia proposta prevede la formulazione di query in cui uno specifico pattern può essere confrontato con un attributo specificato oppure con qualsiasi attributo contenuto nel file. In particolare, il *Basic matching* cerca un pattern in qualsiasi attributo, restituendo così un eventuale superinsieme che soddisfa la query. L'*Advanced matching* mira invece a risolvere una corrispondenza esatta tra coppie specificate di valori e attributi. In questo modo, un utente opportunamente informato sui metadati, può richiedere una corrispondenza con le proprietà implicite del dato, migliorando l'efficacia e l'efficienza del processo di ricerca. Infine, allo scopo di valutare la qualità di un insieme risultante, abbiamo proposto un metodo per confrontare differenti risorse, quando queste sono descritte attraverso i metadati presentati in precedenza. Il nostro approccio è basato su un algoritmo che calcola il grado di similarità tra la richiesta utente e una risorsa trovata, tenendo presente la combinazione di *Basic / Advanced matching* e *Superficial / Deep level* richiesta. In questo modo, il valore restituito dipende dalla conoscenza che l'utente ha del dominio di interesse e quindi acquisisce un significato diverso anche a parità di risorse individuate. Infatti, quando un utente ha una conoscenza approfondita può porre delle query molto dettagliate allo scopo di selezionare solo i dati strettamente necessari.

Diversamente, utenti general-purpose hanno la necessità di porre delle query generiche per ottenere una visione più ampia dei dati che sono ragionevolmente vicini alle loro richieste. Un successivo raffinamento li conduce poi verso i dati di interesse.

Un'ultima osservazione riguarda la struttura e quindi il formato file utilizzato per la memorizzazione dei metadati. L'obiettivo prefissato è garantire l'interoperabilità del sistema, e a tale scopo è stata introdotta una struttura advertisement-like, detta *GeoAdv*, basata su GML (Geography Markup Language) ed utilizzata sia per la memorizzazione che per lo scambio dei dati geografici. GML è un dialetto XML, standard sviluppato originariamente dall'Open Geospatial Consortium (OGC), la cui versione 3.1 è attualmente in fase di sviluppo come standard internazionale ISO 19136 (specifiche di implementazione). L'uso di GML risponde alla necessità di un sistema distribuito che necessita di un linguaggio di comunicazione che sia facile da utilizzare, che abbia una buona componente espressiva, e che soprattutto sia riconosciuto dalla maggior parte delle fonti che accedono a tale sistema. Inoltre, laddove necessario, questo può divenire esso stesso un file di scambio, per dare all'utente interessato una versione preliminare e più leggera della risorsa disponibile.

IL PROGETTO PILOTA: MAP2SHARE VIRTUAL NETWORK

Per sperimentare la nostra proposta, abbiamo testato il motore di ricerca su una architettura P2P ibrida 2-ridondante, dove il network virtuale *Map2Share* (Paolino et al., 2006) è stato realizzato tramite la piattaforma *JXTA*. La figura 1 mostra l'architettura di un singolo peer. Essa è divisa in cinque componenti integrati alla piattaforma *JXTA*, il *Query Translator*, il *Query Resolver*, il *Downloader/Uploader*, il *Visualizer* e l'*Interface*. Inoltre sono presenti due repository, lo *Shared Repository* e lo *Unshared Repository*, che contengono i file gestiti dall'applicazione.

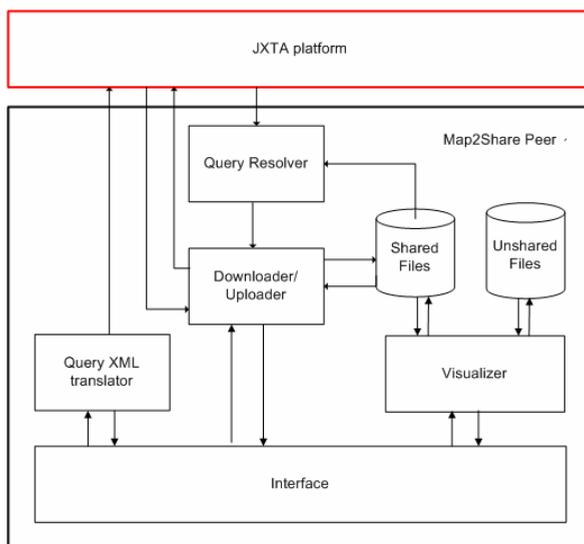


Figura 1: l'architettura di un peer

Il comportamento di ogni componente varia in base al ruolo che il peer riveste, vale a dire *client* o *server*. Nel primo caso, l'utente è interessato a porre delle query e visualizzare il risultato della ricerca. Le query sono composte sull'*Interface*, dove è possibile specificare pattern, attributi, livello e tipo di matching.

Come detto in precedenza, le informazioni riguardanti i metadati e i dati di una risorsa sono organizzati nella struttura *GeoAdv*, che attraverso un insieme di tag, in parte predefiniti ed in parte personalizzati, memorizza tutte le informazioni necessarie a rispondere ai diversi tipi di query.

Una volta che la query è composta, il *Query Translator* la traduce in formato GML e attraverso la piattaforma *JXTA* viene trasmessa al superpeer a cui il peer fa riferimento. Nel frattempo, il peer richiedente si pone in attesa. Ogni volta che un *GeoAdv* arriva, la piattaforma *JXTA* trasmette tale file al modulo *Downloader/Uploader*; che lo instrada verso l'interfaccia per presentare i risultati all'utente. A questo punto l'utente, in base al grado di somiglianza che l'algoritmo ha computato, decide se fare il download della risorsa descritta dal *GeoAdv*. Quando il file è scaricato, l'utente decide se condividerla a sua volta la risorsa, mettendola nello *Shared Repository*, oppure conservarla privatamente come risorsa locale mettendola nello *Unshared Repository*. Quando invece il peer riveste il ruolo di server, il suo comportamento è influenzato solo dalle query che riceve e non da eventuali interazioni con l'utente. In genere, le query vengono ricevute dalla piattaforma *JXTA* e successivamente instradate verso il *Query Resolver* che verifica la somiglianza

tra gli *GeoAdv* condivisi e la query ricevuta.

La figura 2 descrive in dettaglio il comportamento del *Query Resolver* nel momento in cui arrivano delle richieste. Il primo sottomodulo che riceve la query è il *Resolver* ed il suo compito è individuare la combinazione *Basic / Advanced matching* e *Superficial / Deep level* impostata.

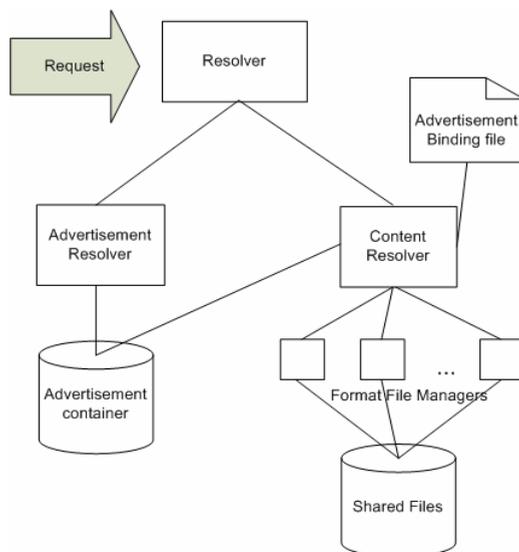


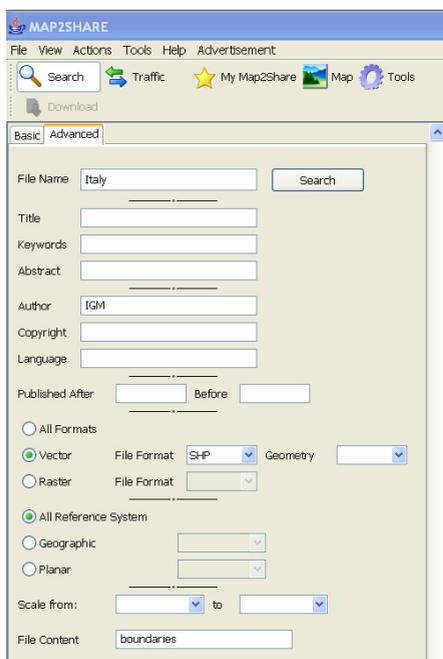
Figura 2: le componenti che gestiscono la richiesta

Nel caso di *Superficial Level*, la richiesta è instradata verso il *GeoAdv Resolver* che esamina gli *GeoAdv* contenuti nel *GeoAdv Container* e restituisce quelli che soddisfano la query. Nel caso di *Deep Level*, la richiesta è trasmessa sia al *GeoAdv Resolver*, sia al *Content Resolver* che verifica il grado di similarità tra la query ed il contenuto dei file dello *Shared Repository*. L'accesso a tali file, a causa della loro natura eterogenea, è gestita attraverso dei moduli specializzati (*File Format Managers*). Al termine di tale operazione gli *GeoAdv* che rappresentano i file che maggiormente soddisfano la query vengono selezionati e ritrasmessi al *Resolver*.

Dal lato del peer richiedente, una volta ricevute delle risposte, esse devono essere mostrate all'utente. Per facilitare questa operazione i risultati sono ordinati in base al grado di somiglianza calcolato tenendo conto della categoria a cui appartiene l'utente.

Le figure 3(a) e 3(b) mostrano alcuni screenshot dell'interazione utente con *Map2Share*. In particolare, la figura 3(a) riproduce un dettaglio della componente *Search* dell'interfaccia per la *Advanced Search*, dove sono settati un insieme di parametri come IGM per *Author* e SHP per *File Format*. La figura 3(b) mostra un dettaglio della *Traffic Area* con evidenziato l'insieme risultato.

Altre componenti importanti di *Map2Share* sono la *MyMap2Share Area* tramite cui è possibile accedere ai repository ove sono memorizzate le risorse, e la *Map Area*, dove, una volta scaricati i file di interesse, è possibile visualizzare le risorse tramite gli usuali strumenti di navigazione e di interrogazione. In tal modo l'utente può verificare immediatamente e all'interno dello stesso ambiente la bontà della risorsa individuata.



(a)



(b)

Figura 3: (a) la Basic Search, (b) un insieme risultato

BIBLIOGRAFIA

- P.A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, *Geographic Information Systems and Science*, 2nd edition, John Wiley & Sons Inc. USA, 2001.
- L. Paolino, M. Sebillio, G. Tortora and G. Vitiello “*Map2Share – A System Exploiting Metadata to Share Geographical Information*”, accettato per la presentazione al 3rd International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS2006), che si terrà a Tucson, USA dal 6 al 9 Novembre 2006, e successiva pubblicazione in Lecture Notes in Computer Science series.
- Dublin Core Metadata Element Set, Version 1.1: Reference Description: <http://dublincore.org/>.
 ISO/TC211, ISO/FDIS 19115 Geographic information -Metadata, <http://www.isotc211.org/>.
- Repertorio Nazionale dei Dati Territoriali - Linee guida per l'applicazione dello standard ISO Mapping between Dublin Core and ISO 19115, “Geographic Information - Metadata” - CWA 14857, November 2003
<http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/cwa/cwa14857.asp>
- OpenGIS® Geography Markup Language (GML) Implementation Specification
<http://www.opengis.net/gml/>