

Integrazione di dati archeologici geografici e non geografici con MAD

Andrea D'Andrea*, Achille Felicetti‡,
Matteo Lorenzini‡, Cinzia Perlingieri*

SOMMARIO. Grazie alla sua versatilità data dalla struttura XML, il linguaggio GML (Geographic Markup Language) dell'Open Geospatial Consortium (OGC) si sta rivelando sempre di più un valido strumento per lo scambio di dati geografici, grafici ed alfanumerici in maniera standard ed interoperabile sia su piattaforme GIS locali che web orientate. Il nostro progetto consiste nell'integrazione del GML con dati non geografici utilizzando l'applicativo MAD (Managing Archaeological Data) sviluppato per la gestione di dataset archeologici strutturati e non strutturati al fine di creare un sistema completamente basato su sintassi XML per la gestione sia dei dati geografici che non geografici.

ABSTRACT. *GIS is an efficient tool for the management of complex geospatial datasets, but geographic information is stored in heterogeneous environments which makes sharing very difficult. To overcome this lack of interoperability, the Open Geospatial Consortium (OGC) has created an XML-based Geographic Markup Language (GML) to provide an XML-based encoding of geo-spatial data and make them portable and flexible enough to be used in different contexts. Data encoded in GML can be integrated with non-spatial data using MAD (Managing Archaeological Data), an application designed to manage structured and unstructured archaeological excavation datasets in order to create complete XML-based systems. This paper will present the GIS extension of MAD enabling the integration with non-spatial excavation information to preserve the native web compliancy of data and the possibility of*

managing unstructured documents (excavation diaries and reports) in a spatial context.

La condivisione delle informazioni geografiche sul web richiede una standardizzazione dei dati geografici che oggi è in via di definizione grazie al contributo di istituzioni quali l'ISO, con gli standard ISO 19100, e l'OGC (Open Geospatial Consortium, <http://www.opengeospatial.org/>) con i loro OpenGIS project e con lo sviluppo dello standard GML.

La struttura dati di una piattaforma GIS consiste nella giustapposizione di differenti tipi di informazione: da una parte ci sono i dati alfanumerici inerenti le informazioni da collocare sulla mappa (generalmente contenuti all'interno di un database), dall'altra ci sono la cartografia di base, che spesso utilizza formati proprietari come DWG (CAD) o SHP (ESRI), e le informazioni relative alla georeferenziazione che sono contenute in un file .tfw, il quale fornisce al file cartografico (raster o vettoriale) le informazioni di collocazione spaziale.

1. L'Open Geospatial Consortium e il GML

Una delle possibili soluzioni al problema dell'integrazione dei diversi dati potrebbe essere l'uso del formato GML utilizzabile sia per applicazioni standalone che per applicazioni web.

Lo standard GML definisce un linguaggio ad oggetti all'interno del quale ogni entità contiene informazioni geografiche e alfanumeriche, per cui:

- ogni entità rappresenta una classe autonoma
- gli oggetti si identificano in base alla loro classe e non al loro codice

Gli oggetti, quindi, sono raggruppati non perché appartenenti alla stessa tabella di dati, ma in base a criteri logici; inoltre un oggetto può contenere altri oggetti, per cui ai dati cartografici potranno essere aggiunti anche dati spaziali e dimensionali, strutturati sugli standard ISO 19107 e ISO 19111, senza dover ricorrere a file esterni.

La sintassi del GML è basata sulla grammatica XML, espressa in XSD, del World Wide Web Consortium, il che costituisce un ulteriore elemento di standardizzazione.

Molti dei principali software di gestione di dati spaziali, sia proprietari che Open Source, sono oggi in grado di gestire il formato GML 2.0, sebbene non supportino ancora lo standard GML 3.0 che permette di gestire molte più informazioni. Tra le principali caratteristiche della versione 3.0 figurano:

- geometrie complesse, non lineari, tridimensionali
- topologia per elementi bidimensionali
- possibilità di visualizzare elementi aventi componenti temporali e/o dinamiche

- utilizzo di sistemi di riferimento e/o unità di misura
- conformità ad ulteriori standard normativi
- struttura dati più adatta alla propria applicazione

I concetti base utilizzati dal GML per la modellazione dei dati geografici sono tratti dall'Abstract Specification dell'OGC (<http://www.opengis.org/techno/abstract.htm>).

GML fornisce numerosi tipi di oggetti utilizzabili per descrivere dati geografici, fra cui entità, sistemi di coordinate, geometria, topologia, tempi e unità di misura;¹ per poter essere rappresentato necessita di uno strumento di rendering capace di interpretarne il contenuto mediante simboli grafici, stili di linee, riempimento di aree e volumi e, spesso, mediante la trasformazione della geometria del dato in presentazione visiva. Per eseguire questa operazione è possibile utilizzare l'Extensible Stylesheet Language Transformation (XSLT), il quale fa uso dell'Extensible Stylesheet Language (XSL) per la trasformazione di un documento XML (ad es. GML) in un altro documento XML (ad es. SVG) secondo regole di trasformazione definite specificatamente. Generalmente tale processo (graphical rendering) trasforma i dati GML in un formato grafico XML.

Nel panorama Open Source esistono vari applicativi in grado di effettuare il rendering: OpenJump, un gis viewer in grado di salvare direttamente i dati con estensione .gml e di aprire direttamente un file .gml senza modificarne la struttura, GRASS, che dalla versione 6.0 permette di salvare il progetto con estensione .gml tramite l'ausilio delle librerie OGR che gestiscono i file vettoriali, e infine il binomio PostgreSQL+PostGIS che, avendo il supporto completo del Simple Features Specifications for SQ definito dall'OGC, permette di vedere i dati vettoriali informatizzati con PostgreSQL direttamente da PostGIS, dando inoltre la possibilità di esportare in formato SVG e GML.

Infine i file GML vengono interpretati nativamente da server cartografici Open Source, ambienti di sviluppo in grado di caricare, gestire e visualizzare sul web un sistema cartografico, quali MapServer con il quale il sistema cartografico viene compilato ad hoc dall'utente con le librerie GDAL e OGR necessarie alla visualizzazione, rispettivamente, dei dati raster e dei dati vettoriali e, quindi, dei file GML.

2. Gestione di dati GML con MAD

MAD è una applicazione web in grado di gestire documenti XML in modo nativo utilizzando XPath/XQuery per la ricerca di specifiche informazioni e XSLT per la presentazione dei risultati.

MAD è in grado di sostituire un tradizionale RDBM, di gestire i dati GML e di effettuarvi query direttamente all'interno del suo database XML-nativo. I frammenti di GML possono essere generati all'istante secondo le richieste dell'utente grazie alle funzioni XQuery per l'implemento delle query geogra-

fiche. L'XQuery processor può anche essere facilmente esteso tramite librerie esterne senza che questo implichi di dover modificare o ricompilare il suo codice sorgente; è in grado di gestire tanto dati geografici, query ed informazioni tratte dal GML, quanto dati non geografici e di presentare i risultati delle query direttamente sul web con programmi volti alla pubblicazione di dati geografici. In breve:

- l'utente crea la propria query semantica e geografica utilizzando l'interfaccia web di MAD e scegliendo la tipologia di presentazione in cui vuole che i risultati gli vengano restituiti;
- la query viene a questo punto elaborata negli archivi XML di MAD su documenti geografici e non geografici;
- il risultato della query viene trasformato e combinato per restituire i dati archeologici relativi alla richiesta e per costruire mappe archeologiche basate sul GML;
- le informazioni geografiche codificate in GML vengono, a questo punto, inviate ad un framework di presentazione per essere mostrate in un browser.

Per restituire i risultati delle query geografiche e creare le mappe e i layer richiesti abbiamo sperimentato l'utilizzo dell'ambiente MapServer, un framework Open Source per la costruzione di applicazioni geografiche orientate al web. MapServer è basato su altri popolari sistemi Open Source o freeware (Shapelib, FreeType, Proj.4, GDAL/OGR) e supporta diverse specifiche web dell'Open Geospatial Consortium (incluso GML), i più utilizzati linguaggi di script e ambienti di sviluppo (PHP, Python, Perl, Ruby, Java, e C#). MapServer è particolarmente versato nel rendering dei dati geografici (mappe, immagini, dati vettoriali) per il web.

Grazie alle funzioni XQuery, MAD è in grado di generare all'istante tutti i dati geografici necessari a MapServer secondo le richieste dell'utente (file MAP, templates, documenti GML, dati raster) eseguendo la query sui documenti GML esistenti ed estraendo le informazioni geografiche a partire dai dati registrati durante gli scavi archeologici (per es. coordinate e livelli di scavo). In quest'ultimo caso i nuovi documenti GML possono essere dinamicamente generati a partire da record archeologici non geografici ed assemblati dall'applicativo per essere mandati al framework map server per la visualizzazione geografica.

Possiamo ormai considerare totale l'integrazione, dal momento che è possibile ricavare i medesimi risultati partendo tanto da documenti geografici che da documenti non geografici.

La totale portabilità del sistema è garantita anche dalla capacità di MAD di trasformare i documenti GML, attraverso fogli di stile XSLT, direttamente in SVG o in KML (in modo da poter essere utilizzati in un ambiente GoogleMaps) ed evitando, in questo modo, di dover implementare un server geografico, attraverso l'utilizzo, al suo posto, delle trasformazioni XML operate da MAD.

Gli oggetti geografici possono essere raggruppati, formattati, trasformati ed assemblati in oggetti già strutturati che, dunque, non necessitano di essere interpretati da nessun ulteriore applicativo. La lista delle funzioni include trasformazioni, clipping path, maschere alpha, filtri, template e estensibilità. La rappresentazione di SVG può essere dinamica ed interattiva per il fatto che il Document Object Model (DOM) di SVG consente la realizzazione diretta di funzioni grafiche interattive attraverso l'utilizzo di eventi Javascript come onmouseover e onclick che possono essere assegnati agli oggetti grafici SVG direttamente da MAD.

3 Conclusioni

L'integrazione dei dati geografici e non geografici contenuti in un solo applicativo e gestiti mediante la medesima interfaccia e la compatibilità web nativa garantita da MAD rendono i dati archeologici pronti per essere interrogati, aggiornati e scambiati sul web, promuovendo lo sviluppo semantico dei servizi web geospaziali.

Lo sviluppo di MAD per la gestione dei dati geografici e non geografici costituisce, in effetti, il primo passo verso la totale e completa implementazione del web semantico geospaziale, un futuro dove il World Wide Web sarà interpretabile dalla macchina e completamente integrata, permettendo la restituzione alle query semantiche di risorse geografiche e non geografiche.

Il Semantic Web sarà in grado di offrire qualcosa che non è mai stato disponibile su vasta scala: l'interoperabilità in grado di legare non solo dati geografici strutturati, come quelli presenti nei database geografici che si trovano al momento on-line, ma anche dati geografici e non geografici non strutturati distribuiti attraverso molteplici archivi web originariamente concepiti dai loro autori per scopi diversi da quelli geospaziali.

Notes

* CISA, Università di Napoli "L'Orientale".

‡ PIN, Università di Firenze.

¹ Il documento di specifica GML 3.0 [GML30] è disponibile sul sito dell'OGC all'indirizzo https://portal.opengeospatial.org/files/?artifact_id=7174.

Riferimenti bibliografici

AA.VV. (2004). *Specifiche per la realizzazione di Database topografici di interesse generale, Intesa Stato Regioni Enti-Locali*. http://www.intesagis.it/specifiche_tecniche.asp.

- Abiteboul, S., P. Buneman e D. Suciu (1999). *Data on the Web: From Relations to Semi-structured Data and XML*. San Francisco: Morgan Kaufmann Publishers.
- Berners-Lee, T. (1998). *SemanticWeb Roadmap, an attempt to give a high-level plan of the architecture of the Semantic WWW*. URL: <http://www.w3.org/DesignIssues/Semantic.html>.
- Chaundri, A. B., A. Rashid e R. Zicari (2003). *XML Data Management - Native XML and XML-Enabled Database Systems*. Addison-Wesley.
- The CIDOC Conceptual Reference Model*. URL: <http://cidoc.ics.forth.gr/>.
- D'Andrea, A., A. Felicetti, S. Hermon et al. (2006). «I linguaggi del W3C e gli strumenti Open Source per la gestione dei dati archeologici». In: *Atti del IWorkshop su Open Source, Free Software e Open Format nei processi di ricerca archeologici*. A cura di Roberto Bagnara e Giancarlo Macchi Jánica. Grosseto.
- D'Andrea, A., A. Felicetti, M. Lorenzini et al. (2007). «Spatial and non-spatial archaeological data integration using MAD, Layers of perception» In: *Proceedings of CAA 2007*. Berlin.
- The EPOCH European Network of Excellence in Open Cultural Heritage*. URL: <http://www.epoch.eu>.
- Extensible Markup Language (XML)*. URL: <http://www.w3.org/XML/>.
- Felicetti, A. (2006). «MAD: Managing Archaeological Data». In: *7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*. Nicosia, Cyprus.
- Geospatial Free and Open Source Software*. URL: <http://www.gfoss.it/>.
- Hermon, S. e F. Niccolucci (2000). «The Impact of Web-shared Knowledge on Archaeological Scientific Research». In: *Proc. of Intl CRIS 2000 Conf*. Helsinki, Finland.
- KML - Guida dell'utente di Google Earth*. URL: http://earth.google.it/userguide/v4/ug%5C_kml.html.
- MapServer*. URL: <http://mapserver.gis.umn.edu/index.html>.
- Open Geospatial Consortium. *Geography Markup Language (GML) Implementation Specification*. URL: https://portal.opengeospatial.org/files/?artifact%5C_id=7174.
- Salminen, A. e F. W. Tompa, cur. (2001). *Requirements for XML Document Database Systems. Vol. Proceedings of the 2001 ACM Symposium on Document engineering*.
- XML Path Language (XPath) Version 1.0. W3C Recommendation 16 November 1999*. URL: <http://www.w3.org/TR/xpath>.
- XQuery 1.0: An XML Query Language W3C Candidate Recommendation 8 June 2006*. URL: <http://www.w3.org/TR/xquery/>.
- XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 November 1999*. URL: <http://www.w3.org/TR/xslt>.
- Zlatanova, S. e D. Prospero (2005). *Large-scale 3D Data Integration: Challenges and Opportunities*. USA: CRC Press Inc.