

MODALITÀ STANDARD PER LA STRUTTURAZIONE DI CARTOGRAFIA NUMERICA PER GIS¹

Andrea SCIANNA^a, Alessio AMMOSCATO^b, Rita CORSALE^b, Rosanna SCIORTINO^b, Benedetto VILLA^b

^a C.N.R. DAST - Dipartimento di Rappresentazione – Università di Palermo, viale delle Scienze c/o Facoltà di Ingegneria, 90128 Palermo – e-mail: scianna@dirap.unipa.it

^b Dipartimento di Rappresentazione – Università di Palermo, viale delle Scienze c/o Facoltà di Ingegneria, 90128 Palermo – e-mail: bevilla@unipa.it

KEY WORDS: cartografia, GIS, standardizzazione, interoperabilità, GML.

RIASSUNTO

Nell'ambito della presente relazione viene presentata una sintesi degli esiti della ricerca, in corso da qualche tempo presso il laboratorio GIS del Dipartimento di Rappresentazione dell'Università di Palermo, finalizzata alla definizione di standard di strutturazione di cartografia numerica per GIS, all'analisi delle caratteristiche di formati di scambio (DXF e GML) e alle modalità di lettura e gestione di questi formati da parte dei principali software, sia proprietari che open; l'obiettivo è quello di verificare l'effettiva interoperabilità del dato geografico su piattaforme software differenti.

ABSTRACT

In the GIS laboratory of "Dipartimento di Rappresentazione" of the University of Palermo some researches are carrying out on the definition of the structure of numeric cartography to use it in Geographical Information Systems, the analysis of the characteristics of exchange formats like DXF and GML, the ways of reading and writing of these formats by main GIS softwares, both proprietary and open source.

In this paper a synthesis of this work is shown. The goal is to verify the real interoperability of geographic data on different software architecture.

INTRODUZIONE

L'utilizzo della cartografia numerica nei GIS impone una riflessione sulle caratteristiche che deve possedere affinché su di essa possano essere eseguite tutte le elaborazioni tipiche di questi sistemi; ci sono infatti delle informazioni che sono necessarie per un sistema e che risultano superflue o addirittura ingestibili in un altro: ad esempio nella rappresentazione di edifici a scala cartografica media e grande il contenuto tridimensionale relativo ai singoli piani è fondamentale per la visualizzazione e le elaborazioni tridimensionali, ma risulta sovrabbondante per la cartografia e spesso non è facilmente gestibile dai software GIS.

La realizzazione di cartografia numerica per un ottimale utilizzo in un GIS comporta la sua strutturazione in un formato specifico, intellegibile da tutti i software esistenti nel mercato, con la necessaria strutturazione topologica e la presenza della componente tridimensionale. Ad oggi la cartografia viene generalmente fornita in alcuni formati CAD (DXF o DWG) o al massimo sotto forma di *shape file*, dal ridotto o inesistente contenuto semantico. Quest'ultimo formato, pur essendo aperto e importabile in tutti i GIS, ha delle limitazioni che sono ravvisabili per esempio nel supporto incompleto della componente tridimensionale e nella ridotta strutturazione topologica (ciò si verifica quando lo strato informativo è generato da software che non consentono una completa strutturazione topologica).

La ricerca si propone di indagare le possibilità di interscambio dei dati in formato GML, analizzando le caratteristiche dei file GML estratti dai diversi software GIS a partire da una base dati grafica comune, importata inizialmente negli stessi software da un formato DXF.

Pertanto, sulla base di quanto detto, al fine di testare l'interoperabilità fra i dati, si deciso di valutare la reale trasferibilità dell'informazione tra software GIS opportunamente strutturati e posti in formato DXF. Esso è attualmente il più semplice e diffuso formato di interscambio, anche se non del tutto adatto alla descrizione dell'informazione geografica.

1. ORIGINE DEL DATO GEOGRAFICO IN FORMATO CAD (DXF)

Le norme che, a livello nazionale e internazionale, sono state fissate con lo scopo di individuare linee guida per la definizione della struttura, delle caratteristiche geometriche e delle relazioni topologiche dell'informazione geografica, sono oggetto di studio ormai da alcuni anni da parte di Enti pubblici e di centri di ricerca. La necessità di più Enti infatti di condividere dati cartografici ha imposto l'esigenza di costituire database centralizzati - per i dati generali (regionali, provinciali) - e database distribuiti - per i dati relativi ad ambiti territoriali più piccoli - capaci di interfacciarsi e consentire a più utenti di eseguire elaborazioni sulla stessa base dati. Se consideriamo facilmente superabile il problema relativo ai permessi di accesso, rimangono comunque da definire alcune questioni

¹ L'attività di ricerca viene svolta sulla base degli indirizzi programmatici del progetto PRIN 2004, dal titolo *Strutture evolute della cartografia numerica per i GIS e l'ambiente WEB*, al quale si sta partecipando.

che riguardano la definizione del contenuto semantico dell'informazione geografica e l'utilizzo di formati di scambio capaci di veicolare velocemente, e attraverso diverse piattaforme, le suddette informazioni.

La definizione di criteri e metodi per la realizzazione di cartografia per GIS deve tenere conto di tutti i tipi di attributo associati all'informazione geografica, in maniera tale che essa risulti idonea per varie tipologie di utilizzo, variabilità di scala e differenti software. In questo contesto rivestono particolare importanza la componente tridimensionale, la correttezza geometrica e la coerenza topologica dei dati (in special modo per le aree).

A partire dal set di dati di esempio (Figura 1), costituito da una serie di informazioni geometriche e topologiche 3D, sono stati testati i software più diffusi, con l'intenzione di verificare in che modo l'uso di un determinato software può incidere sugli attributi geometrici e sulla strutturazione topologica.

I dati, inizialmente archiviati in formato *DXF*, sono stati successivamente importati nei vari software GIS - sia proprietari (ESRI ArcGIS_ArcInfo 8.3/9, Autodesk Map 3D, MapInfo²) che open source (GRASS 6.0) - per verificare la presenza o la perdita delle informazioni e dei vincoli di partenza. Si è quindi proceduto all'esportazione dei dati in formato GML e alla successiva reimportazione degli stessi all'interno delle piattaforme software prescelte; l'obiettivo era quello di verificare l'effettiva capacità del formato GML, prodotto sulla base delle specifiche dell'*Open Geospatial Consortium*, di assumere il ruolo di formato di interscambio del dato geografico.

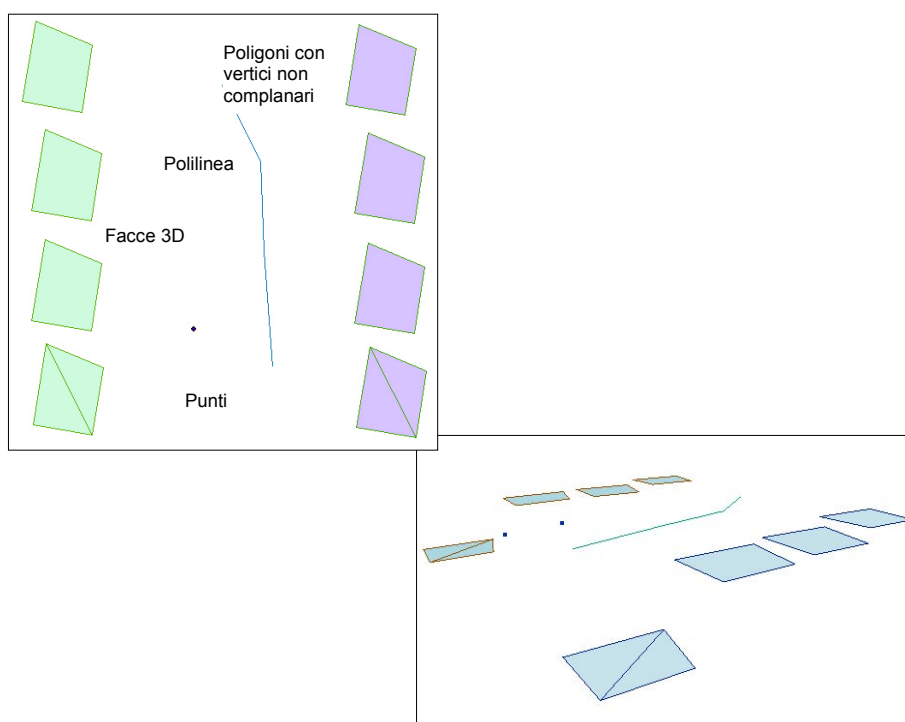


Figura 1: Il set di dati utilizzato. Tutti gli oggetti sono dotati di componente tridimensionale

Gli oggetti che in figura sono rappresentati nella colonna di sinistra sono stati generati in AutodeskMap 3D con il comando 'facce', quelli della colonna di destra sono 'poligoni' (ossia polilinee chiuse), tutti gli oggetti sono dotati di componente tridimensionale e vertici di quota variabile opportunamente definita.

Software testati	Elaborazioni dirette sul file dxf						
	Letture	Solo Importazione	Visualizzazione delle coordinate di tutti i vertici	Calcolo aree proiettate	Calcolo aree effettive	Editing	Cleaning/build topology
Autodesk Map 3D	x		x	x		x	x
ArcGIS 9+ ETGeoWizards	x			x	x		
MapInfo 7.5/7.8		x					
Geomedia Professional 5.0	x		x	x			
GRASS 6.0		x					

Tabella 1: Confronto del comportamento dei software testati nell'importazione di dati in formato DXF

² Tutti i marchi citati in questo documento sono registrati dai legittimi proprietari:

Autodesk Map 3D → Autodesk Inc., <http://www.autodesk.com>

Arcgis – ArcInfo → ESRI, <http://www.esri.com>

Mapinfo → MapInfo Corporation, <http://www.mapinfo.com>

Geomedia Professional → Intergraph Corporation, <http://www.intergraph.com>

È stato verificato che l'importazione del set di dati in formato DXF all'interno dei software GIS (Tabella 1) avviene senza perdita dell'informazione relativa alla terza dimensione, almeno per ciò che riguarda la visualizzazione 3D: la tabella degli attributi associata ad ogni classe di *feature* contiene in genere le informazioni relative alle coordinate del vertice riconosciuto come primo, ed inoltre il valore del perimetro della frontiera tridimensionale e quello dell'area della superficie proiettata sul piano orizzontale.

Per potere importare file in formato DWG o DXF con le versioni più recenti di Grass, per esempio la 6.0 utilizzata in questa sperimentazione, è necessario compilarla utilizzando le librerie "Opendwg". Esse vengono fornite con una speciale licenza che può essere acquisita gratuitamente, previa stipula di un contratto con il quale il contraente si impegna a non ridistribuire il software. Grass 6.0 è stato compilato con tutte le opzioni necessarie, e si è quindi proceduto all'importazione del set di dati. Il software consente comunque l'importazione di file in formato DXF e DWG nelle versioni non superiori alla 2000 di Autocad.

La versione utilizzata per il trasferimento del set di dati non è quindi da sottovalutare nella fase di predisposizione di una cartografia per GIS. L'importazione e la gestione del formato DXF anche all'interno delle versioni 7.5 e 7.8 di MapInfo può essere infatti realizzata solo se i file di origine sono stati salvati nella versione Autocad 12. L'importazione è gestita attraverso due procedure attivabili con le finestre di dialogo 'Import' e 'Universal Translator'; in entrambi i casi il software traduce direttamente i file nel formato proprietario (*tab*).

Le operazioni di *editing* e di *clean* (operazione che consiste nella pulitura dei dati, al fine di eliminare eventuali incoerenze) sul set di dati può essere effettuato solo su AutodeskMap 3D; con gli altri software occorre prima effettuare l'esportazione nel formato proprietario.

2. IL FORMATO DI INTERSCAMBIO: GML

Già da alcuni anni l'*Open Geospatial Consortium* ha sviluppato il linguaggio GML, orientato alla standardizzazione e la strutturazione di dati GIS, finalizzato al trasferimento ed all'uso dell'informazione geografica fra diverse piattaforme e tramite internet. Il GML, in quanto sistema di codifica XML specifico per l'informazione geografica, può essere visualizzato anche con un editor di testo, con il conseguente vantaggio di lettura e scrittura diretta delle informazioni, e può descrivere oggetti senza dovere ricorrere a formati proprietari. L'utilizzo del GML rappresenta un passo avanti verso l'interoperabilità: a partire dal formato DXF, utilizzato su sistemi CAD, e convertendo la banca dati in GML dovrebbe essere possibile consentire un facile accesso a informazioni geografiche contenute in banche dati geografiche distribuite che utilizzano formati, come l'SVG, per cui sono oggi disponibili dei *plugins* per la visualizzazione su *browser* internet.

Le entità presenti in uno schema GML sono descritte seguendo le istruzioni dettate dall'*Open Geospatial Consortium*, relativamente a strutturazione delle informazioni, tipi di oggetti geografici, metadati, definizione delle relazioni geometriche e topologiche delle entità.

Allo stato attuale la versione più avanzata del linguaggio GML è la 3.1, che però mantiene ancora una serie di costrutti che garantiscono la retrocompatibilità con la versione 2 di GML; l'unico software che gestisce il GML 3 è Geomedia Professional 5.2.

Tra le due versioni di GML vi sono però numerose altre differenze, fra le quali le principali consistono nell'introduzione in GML 3 di:

- geometrie complesse, non lineari e tridimensionali;
- topologia per elementi bidimensionali;
- possibilità di rappresentare elementi aventi componenti temporali e/o dinamiche;
- utilizzo di sistemi di riferimento e sistemi di unità di misura;
- conformità ad ulteriori standard normativi.

Anche la definizione degli "Application Schemas" è stata notevolmente modificata e allargata, in funzione delle novità sopra esposte. All'interno di un documento XML è possibile individuare un albero di elementi, ciascuno dei quali ha un "nome di tipo" (tag) ed un insieme di attributi, composti da un nome ed un valore; ai tag è possibile attribuire qualsiasi nome che non contenga spazi, ma è buona norma utilizzare nomi che diano dei chiari riferimenti all'informazione che forniscono.

Le prime righe del documento fungono da intestazione: la prima contiene istruzioni di elaborazione, che indicano la versione di XML che si sta usando ed il tipo di codifica dei caratteri; le successive forniscono la descrizione del "namespace", ossia di una collezione di nomi (identificati da un riferimento URI) che sono usati in un documento. In particolare in un file GML sono indicati gli indirizzi cui sono associati i prefissi adoperati per definire le *Feature Collection* e gli attributi (geometrici e non) delle stesse, ed inoltre anche quelli che localizzano gli schemi.

Software testati	Letture	Solo Importazione	Visualizzazione delle coordinate di tutti i vertici	Elaborazioni dirette sul file GML2	Supporto GML3
Autodesk Map 3D		x	x		
ArcGIS9		x			
MapInfo 7.5/7.8		Solo MasterMap GML			
Geomedia Professional 5.2	x		x		x
GRASS 6.0 con librerie aggiuntive		x			

Tabella 2: Confronto del comportamento dei software testati nella gestione del formato GML

I test condotti sui software hanno evidenziato l'impossibilità di effettuare una vera e propria lettura dinamica di un file GML: l'operazione consentita è quella di importazione attraverso la trasformazione nel formato proprietario; fa eccezione Geomedia Professional, che gestisce tutti i tipi di file (GML, CAD, shapefile, tab, ecc.) creando con essi una connessione in sola lettura. Questo

comportamento costituisce un limite all'interoperabilità, che è stata già realizzata fra alcuni software CAD³: al fine di eseguire operazioni di *editing*, è necessario infatti importare il file in GML nel software utilizzato, trasformandolo nel formato proprietario, e poi riesportarlo in GML.

Uno degli obiettivi primari del GML è quello della standardizzazione e della conseguente realizzazione dell'interoperabilità, per questa ragione gli schemi di tutti i file prodotti fanno riferimento alle norme (*GML Implementation Specification*) definite da *Open Geospatial Consortium* per GML2.

Ulteriore ostacolo all'interoperabilità, oltre alla impossibile utilizzazione dinamica, è rappresentato dal fatto che ciascuna casa produttrice di software GIS realizza file GML che, seppure 'corretti' perché rispondenti alle norme, utilizzano tutte le possibilità di definizione libera consentita dalla norma stessa. Accade quindi che la lettura incrociata dei file GML da parte dei software testati risulti spesso complicata, se non impossibile. Al fine di comprendere meglio tali problematiche sono stati presi in esame i file GML prodotti dai vari software, esaminandone le differenze di struttura, di sintassi, delle proprietà (geometriche e non) associate agli oggetti.

In particolare è stato osservato che a partire dalle prime righe di questi file, dove sono allocate le descrizioni dei namespace (dominio di esistenza dei nomi), vi sono delle indicazioni differenti: questo comporta già un "decadimento di compatibilità".

In genere insieme ai file GML i software creano file di schema con estensione *xsd*, il cui posizionamento viene indicato nell'intestazione degli stessi file GML. ArcGIS9, nell'esportazione in modalità FIXED, crea file GML2 in cui l'allocatione dello schema non tiene conto del nome del file creato, ma di un percorso generale che fa riferimento al sito FME (*Feature Manipulation Engine*), che è produttore del motore del software di esportazione. Dai confronti effettuati è stato rilevato che questo tipo di file GML viene importato facilmente (completo del contenuto 3D) all'interno di AutodeskMap 3D, con cui ESRI ha una maggiore compatibilità.

Esistono però diverse possibilità, utilizzate in svariate combinazioni, di definizione:

- dell'allocatione del file-*schema* specifico, tramite l'indirizzo internet e/o la posizione locale;
- delle coordinate dello spazio che contiene le istanze della classe e dei vertici delle istanze;
- del sistema di riferimento;
- degli attributi non geometrici.

Occorre in ultimo precisare che:

- affinché GRASS 6.0 possa interagire con file GML (limitatamente alla versione 2.0 del GML) è necessario compilarlo opportunamente, ovvero ricompilare aggiungendo le librerie 'gdal' con il supporto 'exceres'. Per la sperimentazione si è quindi provveduto alla configurazione di GRASS includendo le librerie gdal, compilate come sopra indicato, e le librerie opendwg (che non sono "open" in senso stretto, come detto prima, poiché è permesso solo un uso personale dopo la firma di un apposito contratto);
- per l'utilizzo del GML attraverso Geomedia Professional 5.2 è necessario installare alcuni plugins e delle librerie aggiuntive (GML Data server, GML export service, GML export command), tutti scaricabili gratuitamente dal sito del produttore.

In nessuno dei software viene utilizzata una strutturazione dei dati che possa fungere da supporto alla costruzione della topologia, in quanto ciascuna istanza di una classe viene definita senza alcuna relazione con le altre; questo tipo di strutturazione impedisce, ad esempio, di definire un poligono in funzione di archi. Ciò avviene poiché i software GIS, pur contenendo le informazioni topologiche, non sono in grado di trasportarle nel GML nella fase di esportazione. Questo genere di problemi verrà superato nel momento in cui i software supporteranno pienamente le caratteristiche della versione 3 del GML.

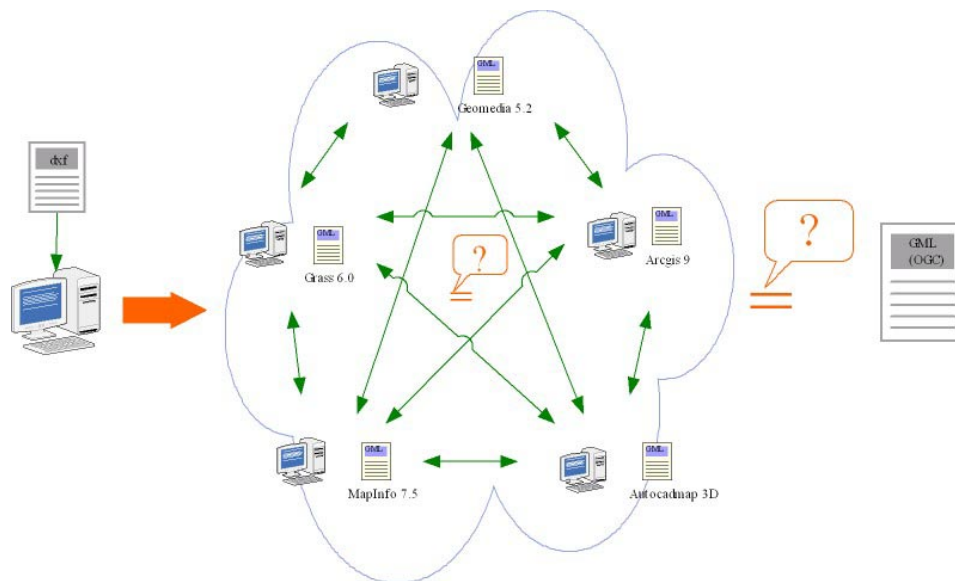


Figura 2: Schematizzazione ideale dell'interoperabilità fra i dati

³ Microstation 8, software CAD, apre i file in formato DXF e DWG, consente di lavorare sugli stessi e di salvarli nello stesso formato, per essere aperti successivamente con Autocad o con Rhinoceros.

3. CENNI SULLA STRUTTURAZIONE DEI DATI

L'interoperabilità non passa solo attraverso le caratteristiche del formato di scambio, ma anche attraverso una struttura condivisa delle informazioni e una definizione univoca, e indipendente dalle piattaforme utilizzate, delle entità geometriche.

In quest'ottica una parte del lavoro è stata finalizzata alla definizione della struttura dei dati cartografici; da un'analisi di proposte di cartografia numerica basata sulle raccomandazioni della Commissione Geodetica Italiana, si è proceduto in una prima fase a definire una struttura ad albero, tipica dei sistemi CAD ed esplicitata tramite layer. Riconosciuti i limiti di tale impostazione per l'utilizzo in ambiente GIS, è stato effettuato uno studio per la definizione della struttura della cartografia finalizzata alla sua allocazione in un geodatabase (definizione delle classi e loro associazioni, definizione degli attributi, definizione dei vincoli, ...). Parallelamente si sta affrontando la problematica della gestione e trasposizione di strutture di dati cartografici, tramite opportuni software (Rational Rose, Visio) che utilizzano il linguaggio UML per generare un database spaziale a partire dallo schema creato dagli stessi software. Va osservato che allo stato attuale questi schemi sono stati creati a partire da modelli relativi a software GIS specifici, in particolare è stato utilizzato l'ArcInfo UML Model e quindi il database generato può essere utilizzato (per il popolamento) solo dai software della stessa casa produttrice.

Sono in corso ulteriori studi finalizzati a verificare l'utilizzo di modelli diversi che non abbiano relazioni con software specifici; lo scopo è quello di incrementare l'interoperabilità anche in fase di progettazione e realizzazione della struttura dati.

CONCLUSIONI

L'utilizzo del formato DXF come formato di interscambio pone svariati problemi, sia a livello topologico che geometrico.

I diversi software, leggendo (Autodesk Map 3D, ArcGIS, Geomedia) o importando (MapInfo, GRASS) lo stesso file DXF, mostrano comportamenti eterogenei, non riuscendo talvolta a individuare correttamente la geometria di un oggetto e le semplici relazioni topologiche. Inoltre, due oggetti aventi le stesse caratteristiche geometriche, ma realizzati uno come faccia o l'altro come poligono (polilinea chiusa), vengono spesso identificati dallo stesso software in maniera differente. Ciò vuol dire, per esempio, che alcuni nodi di questi oggetti vengono identificati come semplici vertici. La conseguenza è, in questo caso, una diversa attribuzione della quota, essendo questa identificata da tutti i software come quella del primo vertice.

Per ciò che riguarda la gestione del contenuto topologico degli oggetti: punti, linee e polilinee tridimensionali, sono gestiti correttamente dai software e quindi è possibile applicare ad essi le funzioni di 'clean' e 'build topology', dopo opportuno trasferimento nel formato proprietario. Lo stesso non vale per le superfici nello spazio: è facile verificare che i software testati le gestiscono come se fossero idealmente "proiettate" su un piano, e che anche l'area che di esse viene calcolata è in realtà quella della loro proiezione sul piano orizzontale.

Alcuni software hanno la possibilità, attraverso opportune estensioni (Et-Geowizard per ArcGis), di superare almeno in parte queste limitazioni, dal momento che consentono, per esempio, la visualizzazione delle quote dei vertici e il calcolo delle aree effettive.

Per quanto attiene alla sperimentazione sul GML, è stato osservato che, nonostante le più importanti società produttrici di software affermino di fare riferimento agli standard definiti nell'ambito di *Open Geospatial Consortium*, non si è ancora raggiunta una standardizzazione dei formati a livello operativo. Alcuni case produttrici di software tentano addirittura di affermare un GML non standard, avente delle caratteristiche che lo rendono utilizzabile solo con il proprio prodotto. Ulteriore incertezza è data dall'attuale transizione fra le diverse versioni del GML che, malgrado gli sforzi fatti nella direzione della retrocompatibilità, comporta spesso dei malfunzionamenti: la maggior parte dei software risulta essere pienamente compatibile con la versione 2 del GML. Nel corso della sperimentazione sono stati riscontrati anche problemi di compatibilità su uno stesso software, esportando e reimportando in formato GML.

Rimane comunque buona la valutazione delle potenzialità di un tale strumento, in quanto capace di superare molti dei problemi legati all'interoperabilità.

Bibliografia

Bibliografia da libri

- AA. VV. (2004). Implementazione di un database geografico secondo le Specifiche di contenuto per la realizzazione dei Data Base Topografici di interesse generale - Versione Aerofotogrammetrica ed analisi delle problematiche legate alla convivenza nello stesso DataBase di dati acquisiti con precisioni diverse (DB multiscala).
- Erik T. Ray (2001). *Learning XML* O'Reilly
- AA. VV. (1999). Norme tecniche per la realizzazione di cartografia numerica alle scale nominali 1:1000 e 1:2000, Milano.
- Atzeni P., Ceri S., Paraboschi S., Torlone R. (1999). Basi di dati. McGraw-Hill, Milano.
- Commissione Geodetica Italiana (1976). La formazione di cartografie generali a grande scala (1:2000 e 1:1000). Le strade, Milano.
- Commissione Geodetica Italiana (1973). Norme proposte per la formazione di carte tecniche alle scale 1:5000 e 1:10000. I.G.M. Firenze.

Bibliografia da web site

- GML 3.00 (2003) OGC 02-023r4 <http://www.opengeospatial.org/specs/?page=baseline>
- ArcGIS - ArcMap (2004). <http://www.esri.com/software/arcgis/arcmap/index.html>.
- ESRI (2004). Open Published Data Format, <http://www.esri.com/software/opengis/openpdf.html>.
- Intergraph (2004) Geomedia Professional, <http://imsgs.intergraph.com/geomedia/default.asp>
- Namespaces in XML <http://www.w3.org/TR/1999/REC-xml-names-19990114/>
- The schema Element http://www.w3schools.com/schema/el_schema.asp (accessed 2005 May 26)
- XML Schema Instance Attributes <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/xmlsdk/html/822217a9-5282-41e4-ac9e-21fb0f5959bd.asp> (accessed 2005 May 26.)

FME (2005). <http://www.safe.com/products/fme/>.
Piazza Bonati L., Fortunati L., Fresta G. (2005). SVG Explorer of GML DATA
<http://webgisserver.cnuce.cnr.it/centroSi/progetti/SvgViewer/svgViewer.html> (accessed 2005 May 24)