

L'adozione del GML nella tecnologia ESRI: gestire la complessità del GML 3 con il GML Simple Features Model

Gianni CAMPANILE (*)

(*) ESRI Italia S.p.a., Via Tiburtina, 755 – 00159 Roma, tel: 06 40696245 fax: 06 40696333,
email: gcampanile@esriitalia.it

RIASSUNTO

Il GML è una specifica XML dell'OGC per descrivere dati geografici e loro caratteristiche, quali attributi, topologia etc. La versione attuale del GML, la 3.1.1, è una specifica molto completa ma al contempo troppo complessa per la maggior parte dei casi reali. Attraverso l'uso di profili, cioè di sottoinsiemi dei tipi e delle regole, è possibile ridurre e gestire efficacemente tale complessità. La specifica GML-SF di OGC risponde a questi requisiti, e rappresenta un valido compromesso tra completezza e semplicità d'uso. ESRI, membro principale di OGC, è uno dei proponenti e sostenitori di tale specifica e uno dei primi produttori GIS ad adottarla nei propri prodotti.

ABSTRACT

GML is an OGC specification for how to use XML to describe geospatial features and their characteristics, like attributes, geometry, topology etc. The actual version, GML 3.1.1, is very rich and complete but at the same time is too complex for the majority of real world use cases. Using Profiles, i.e. a well defined subset of types and rules, it is possible to reduce such complexity. The GML-SF specification from OGC is a GML Profile that is sufficiently complete but at the same time reduces dramatically the effort needed for implementation. ESRI, a principle member of OGC, is one of the contributors of GML-SF and one of the first GIS vendors to adopt it in his products.

KEYWORDS: XML, XSD, GML, Simple Features, Schema, Profile, ISO, OGC, SQL/SF, WFS

INTRODUZIONE

Il GML (*Geography Markup Language*) ha le sue radici nei vari dialetti nati per la necessità di scambiare dati di natura geografica:

- US Geological Survey's Digital Line Graph (DLG)
- US Spatial Data Transfer Standard (SDTS)
- Canada Spatial Archive and Interchange Format (SAIF)
- ...e molti altri.

Il problema da superare era, ed è, quello di formattare i dati in modo indipendente dal software di uno specifico venditore GIS. Dal punto di vista pratico, molto spesso la soluzione adottata era quella di aderire ad uno standard industriale diffuso, scegliendo un compromesso tra il costo interno di conversione dati, la completezza dello standard e la sua diffusione. Ovviamente questo

portava ad un proliferare di file aggiuntivi, secondo formati del tutto arbitrari, la cui semantica era conosciuta solamente al produttore e al consumatore.

Attualmente vi sono diverse tecniche, a volte chiamate *design patterns*, per lo scambio dei dati spaziali fra sistemi eterogenei:

- *File batch*: i dati geografici sono formattati in un formato concordato e poi trasmessi in modalità batch o letti tramite download;
- *API (Application Programming Interface)*: i dati sono letti attraverso l'uso di un insieme di funzioni messe a disposizione dal sistema GIS. La semantica e la sintassi delle funzioni sono dipendenti dal software stesso.
- *Web Service*: i dati sono acceduti e letti attraverso una serie di protocolli standard e, se necessario, attraverso Internet. La semantica e sintassi delle funzioni viene esplicitata, ma *non* i dati utilizzati.

Queste tecniche permettono un accesso ai dati tramite diverse modalità, ma ancora non forniscono una soluzione al *formato* dei dati che vengono scambiati.

Il *GML (Geography Markup Language* [3]) è uno standard basato su XML sviluppato dall'OGC (Open Geospatial Consortium [4]). Le specifiche GML definiscono uno *Schema*, cioè una specifica per come utilizzare il linguaggio XML per la descrizione di features, della loro geometria, grid, dati temporali, sistemi di coordinate e ogni altro necessario per un GIS. L'uso di XML permette l'utilizzo di tale formato per tutte le tecniche di scambio di dati geografici descritte in precedenza.

GESTIRE LA COMPLESSITA' DEL GML CON IL PROFILO GML-SF

La versione attuale pubblicata del GML è la 3.1.1 ([3]) (la versione 3.2 è in corso di definizione). Questa specifica di circa 600 pagine ha introdotto molte novità rispetto alla versione 2.x, quali la gestione delle coordinate temporali, della topologia, dell'unità di misura, di geometrie non lineari che la rendono sicuramente molto completa ma anche complessa da gestire soprattutto in caso di processamento "on-the-fly", piuttosto che di elaborazione batch.

Uno studio comparativo fra 63 diversi tipi di schema XML effettuato da Stan Kitsis di Microsoft ([1]), mostra come il GML sia considerato uno schema fra i più complessi, come risulta dalle tabelle in *Figura 1*, che considera il numero di linee di codice e in *Figura 2* che considera il numero di tipi complessi definiti.

Numero di linee di codice	Categoria	Numero di schemi
0 – 100	Mini	0
100 – 1,000	Small	12
1,000 – 10,000	Medium	24
10,000 – 100,000	Large	23
GML 3.1.1: 10,291 linee di codice		
100,000 – ...	Huge	4

Figura 1 - Suddivisione di schemi XML secondo il numero di linee di codice

Numero di tipi complessi	Categoria	Numero di schemi
0 – 32	Mini	13
32 – 100	Small	12
100 – 256	Medium	14
256 – 1,000	Large	12
GML 3.1.1: 287 tipi complessi		
1,000 – ...	Huge	12

Figura 2 - Suddivisione di schemi XML secondo i tipi complessi definiti

Al di là di una valutazione quantitativa, esperienze effettuate in progetti reali mostrano come, molto spesso, non sia necessario disporre della potenza espressiva del GML 3.1.1: è quindi possibile e auspicabile utilizzare un formalismo meno complesso.

Le strade da seguire per gestire tale complessità sono principalmente 2:

- *Application Schema*
- *Profile*

Uno *Schema* è in generale una descrizione della struttura e delle regole che un documento XML deve soddisfare per essere conforme ad un certo tipo, mentre un *Profile* è un sottoinsieme di elementi della specifica, con l'aggiunta di restrizioni sulle regole di dipendenza.

Entrambe le strade portano ad una semplificazione del GML generale, ma gli obiettivi e i vantaggi/svantaggi sono diversi, come illustrato nella seguente tabella riassuntiva:

Specificata	Definito da	Descrive...	Vantaggi	Svantaggi
GML Core Schema	OGC	<ul style="list-style-type: none"> • Features, coverage, misure, etc. Geometria, sistemi di riferimento; • Metadata, altri costrutti geospaziali ; • Regole per estendere gli schemi GML; 	<ul style="list-style-type: none"> • Completo • Standard 	<ul style="list-style-type: none"> • Complesso
Application Schemas	Utenti, casi d'uso, IT	<ul style="list-style-type: none"> • Modelli per domini applicative, costruiti sulla base del Core Schema; • Ad.es. schema per i trasporti; schema per le Reti Idriche; schema per il terreno; 	<ul style="list-style-type: none"> • Meno complesso 	<ul style="list-style-type: none"> • Meno noto • Parser specifici per ogni schema
GML Profiles	OGC	<ul style="list-style-type: none"> • Sottoinsieme di elementi del GML; • Restrizioni sull'uso degli elementi; 	<ul style="list-style-type: none"> • Meno complesso • Generico 	<ul style="list-style-type: none"> • Non copre tutti i casi

Figura 3 - Tabella riassuntiva di schemi e profili GML

Come si vede dalla tabella non esiste una soluzione unica ottimale; d'altra parte se si riuscisse a definire un profilo sufficientemente ampio da coprire la maggior parte dei casi d'uso, sarebbe

probabilmente una valida alternativa al GML completo, riducendo drasticamente la curva di apprendimento e la complessità dei parser che devono interpretare il file.

E' con questo scopo che è stato definito da OGC il profilo *GML-SF (GML Simple Features Profile)*([2]) cioè un profilo di GML adatto a trattare le Simple Features. Questo modello, pubblicato ad Aprile 2006 nella versione 1.0.0, è il completamento di un lavoro iniziato nel 2003 basato sulla definizione delle Simple Features (cfr. [9] e [5]).

Il modello Simple Features, adottato da OGC e da ISO, definisce:

- Il modello dati
- Il linguaggio di accesso e manipolazione dati
- Geometria 2D per punti, linee e poligoni
- Formati dati "Well-Known" per geometria e riferimenti spaziali.

La definizione del GML-SF va a ricoprire il modello Simple-Features, costituendo un sottoinsieme valido, usabile e relativamente semplice del GML. Consideriamo ad esempio il seguente frammento di specifica GML-SF per la definizione di un feature type:

```
<xsd:element      name="[FeatureTypeName]"
                  type="[prefix]:[FeatureTypeName]Type"
                  substitutionGroup="gml:_Feature" />
```

.....

E' possibile notare come solo la classe *_Feature* e le sue sottoclassi (*_Coverage*, *Observation*) possono essere utilizzate in un application schema compatibile con questa specifica. Questo esclude topologia, dati temporali, e altro.

Oltre a trattare un sottoinsieme di oggetti, il GML-SF ammette per le proprietà (attributi) di una feature solo i seguenti tipi base:

- Integers
- Reals
- Character strings
- Date
- Boolean
- Binary data
- URI

Dalla lista è esclusa la possibilità che un attributo possa essere un'altra feature, una restrizione che semplifica notevolmente lo schema dati. Un'altra semplificazione notevole è l'impossibilità di creare application schema "dinamici", cioè il GML-SF ha uno schema fisso, non estendibile. Queste caratteristiche costituiscono sì un limite della specifica ma rendono sicuramente più efficienti i programmi ed in particolare i server *WFS (Web Feature Service)*.

Nell'ambito del GML-SF vengono definiti 3 livelli di specifica chiamati SF-0, SF-1 e SF-2, di complessità crescente. Mentre il livello SF-0 impone delle limitazioni sull'insieme dei tipi utilizzabili per le proprietà di una feature e sulla loro cardinalità, negli altri livelli queste limitazioni sono via via superate fino ad arrivare al livello SF-2 che corrisponde alla specifica OGC per le simple features (cfr. [2] e [9]).

L'ADOZIONE DEL GML 3 e GML-SF NELLA TECNOLOGIA ESRI

ESRI è sempre stato un attore principale nella definizione di standard internazionali, essendo presente ai tavoli dei principali enti internazionali (ISO, OGC, OASIS, FGDC, W3C, etc.) e proponendo essa stessa formati poi divenuti standard “de facto”, quali ad esempio il formato Shapefile. Inoltre l’attenzione di ESRI al tema dell’interoperabilità risulta evidente nel continuo adeguamento della tecnologia ArcGIS, per mantenere l’aderenza agli standard dettati dall’Information Technology e a quelli più specifici del settore del GIS. Infatti le principali caratteristiche della tecnologia ArcGIS sono:

(in ambito IT)

- multi piattaforma (Windows, Linux, Unix,...);
- gestione della componente geografica su DBMS eterogenei (MS Access, DB2, INFORMIX, Oracle, PostGRES e SQLServer), sia in formato ESRI che nel formato spaziale nativo del DBMS;
- personalizzazione applicazioni con l’uso di ambienti di sviluppo standard (VBA, C++, Java, .NET, Python,...);
- realizzazioni di architetture GIS di tipo DeskTop, Client/Server, Internet/Intranet, SOA (Service-Oriented Architecture);

(in ambito GIS)

- oltre 60 i formati dati (vettoriali/raster) accessibili direttamente o tramite appositi convertitori nei prodotti base;
- supporto degli standard WMS (Web Map Service), WFS (Web Feature Service), CSW (Catalog Service);
- pubblicazione in XML del formato del GeoDatabase (il nuovo **modello** dati ESRI);
- gestione dei Metadata negli standard FGDC, ISO TC19115 e ISO TC19139 (dalla 9.2)

Per quanto riguarda l’Open Geospatial Consortium ESRI è un membro principale e partecipa attivamente con diversi ruoli ai vari comitati tecnici e decisionali e ai programmi di interoperabilità e standards. Nell’ambito della definizione di dati e metadati presiede o co-presiede nei seguenti Working Groups:

- Metadata Working Group (WG)
- Location Services Working Group (WG)
- **Web Feature Service WG**
- **Simple Features Revision Working Group (RWG)**
- **GML Simple Features RWG**

In aggiunta ESRI è uno dei proponenti della specifica GML-SF, ed è per questo che tale specifica si trova già integrata nella propria suite ArcGIS. Già dalla versione 9.0, attraverso le estensioni *ArcGIS Data Interoperability* e *ArcIMS Data Delivery* è possibile leggere e scrivere dati in formato GML, mentre nella prossima versione ArcGIS 9.2 è implementato anche il formato GML-SF.

La tabella seguente mostra le relazioni tra le versioni di ArcGIS e il formato GML:

Specification	Provider	Consumer
SQL Simple Features 1.1	ArcSDE® ArcGIS	ArcGIS

Geography Markup Language (GML) 2.x e 3.1.1 GML-SF 1.0.0	ArcIMS WFS Connector—9.1 ArcIMS Data Delivery Extension—9.x ArcIMS—9.2 ArcGIS 9.2 Server	ArcGIS Data Interoperability Extension—9.x ArcGIS Data Interoperability Extension—9.2 (senza licenza) ArcGIS 9.2
--	---	--

Figura 4 - Implementazione del GML nella suite ArcGIS

L'estensione ArcGIS Data Interoperability, sviluppata da ESRI in collaborazione con Safe Software, leader mondiale nel trattamento e conversione di dati geografici, permette la lettura e scrittura diretta di ulteriori 65 differenti formati dati, già pronti e altri tramite il WorkBench, uno specifico ambiente che consente la realizzazione di connessioni a formati dati personalizzati:

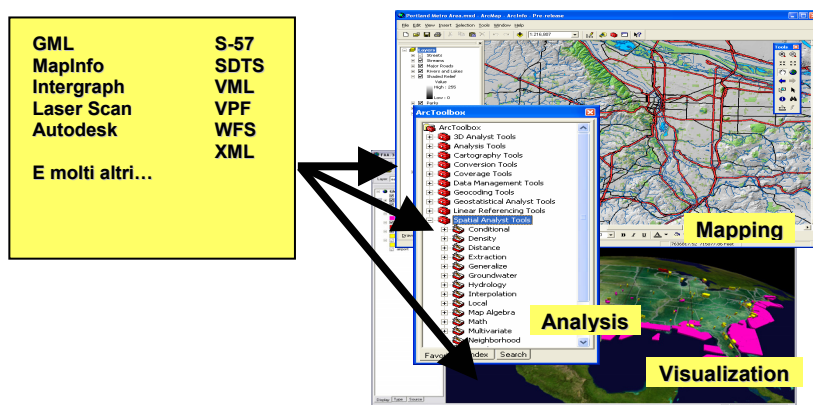


Figura 5 - L'estensione Data Interoperability in ArcGIS

Il supporto di ESRI per l'interoperabilità è continuo e, in particolare per il GML, si concretizza in:

- Definizione di application schema basati su esperienze reali;
- Proposte di nuovi profili per includere più casi possibile senza andare a detrimento della semplicità
- Implementazione dei nuovi standard nei propri prodotti

CONCLUSIONI

L'adozione del profilo GML-SF porterà ad una semplificazione nell'interoperabilità, intesa sia come scambio dati che come capacità di interazione da parte di sistemi GIS con formati dati eterogenei.

In questa continua evoluzione ESRI è sicuramente un attore principale, e la sua tecnologia, basata su anni di esperienza, insieme alla sua attiva partecipazione ai principali tavoli di definizione di standard e tecnologie, è allineata con gli standard più evoluti e tecnicamente più validi.

BIBLIOGRAFIA

- [1] Ralf Lammel, Stan Kitsis, Dave Remy: Analysis of XML Schema usage
<http://homepages.cwi.nl/~ralf/xml05/paper.pdf>
- [2] GML 3.1.1 Simple features Profile 1.0.0
http://portal.opengeospatial.org/files/?artifact_id=15201
- [3] OpenGIS Geography Markup Language specification 3.1.1
http://portal.opengeospatial.org/files/?artifact_id=4700
- [4] Open Geospatial Consortium, <http://www.opengeospatial.org>
- [5] ISO 19125-1: Geographic information – Simple feature access - Part 1: Common architecture
- [6] ISO 19125-2: Geographic information – Simple feature access - Part 2: SQL Option
- [7] ISO/DIS 19136: Geographic information - Geography Markup Language
- [8] ESRI Supported ISO and OGC standards
<http://www.esri.com/library/whitepapers/pdfs/supported-ogc-iso-standards.pdf>
- [9] OGC Simple feature access part 1 ver. 1.1.0
http://portal.opengeospatial.org/files/?artifact_id=13227

