

# **Ontologie per l'Informazione Geografica: prospettive e implementazione**

Roberto FRESCO (\*), Roberto DELLA MAGGIORE (\*\*)

(\*) Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI – CNR),  
Via Moruzzi 1, 56124 Pisa, tel. 0503153075, fax 0503152040, Roberto.Fresco@isti.cnr.it

(\*\*) Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI – CNR),  
Via Moruzzi 1, 56124 Pisa, tel. 0503152944, fax 0503152040, Roberto.dellaMaggiore@isti.cnr.it

## **RIASSUNTO**

*L'Informazione Geografica presenta aspetti di modellazione di features e di fenomeni, nello spazio, che determinano di fatto l'implementazione di un sistema e l'interazione tra diversi sistemi GIS. Le necessità di interoperabilità di dati, strutture e processi in ambito geografico coinvolgono aspetti non soltanto puramente tecnici ma anche semantici, legati alla rappresentazione di una conoscenza condivisa di entità geografiche. Questo articolo descrive prospettive e metodologie per la formalizzazione di semantica per l'informazione geografica (ontologie) utilizzabile in modo automatico attraverso interazioni machine to machine nel contesto distribuito del Web.*

## **ABSTRACT**

*Some aspects of Geographical Information, relating to spatial features modelling (as well as to phenomena modelling), heavily affect the systems implementation and the interaction process between different GIS systems. The needs of interoperability of data, structures and processes in a geographical environment involve not only technical issues, but semantic ones too. For geographical entities, semantic aspects are deeply linked to the definition of a shared knowledge representation. This work sketches the perspectives and the methodologies for the formalization of semantics (ontology) for geographical information in order to make geographical information automatically sharable by machine to machine interactions on the Web.*

**KEYWORDS:** *Ontology, Semantic Web, knowledge representation, RDF, OWL.*

## **INTRODUZIONE**

L'utilizzo di Internet e del World Wide Web come piattaforma distribuita ed universale di informazione ha modificato ed esteso anche la fruizione e la diffusione di Informazione Geografica. Il GIS su Web non è solamente uno strumento di condivisione e recupero di dati spaziali, ma sta diventando una piattaforma effettiva di elaborazione distribuita che mette in interazione sistemi GIS anche molto eterogenei tra loro, principalmente attraverso il paradigma orientato ai servizi e l'insieme delle tecnologie XML.

Da questa situazione emergono problemi di interoperabilità di due tipologie: *tecnologica* in quanto i sistemi GIS producono diversi tipi e formati di dati e la loro condivisione è resa complessa dalla necessità di strategie di conversione ed armonizzazione dei dati strutturati che sono codificati in forme eterogenee; *semantica* perché la modellazione di features spaziali avviene solitamente secondo schemi e standard propri di ogni organizzazione. L'esposizione di questi modelli su Internet comporta problemi di interpretazione non ambigua.

Il grado di matching tra ciò che un utente si aspetta di trovare con una propria richiesta e ciò che viene effettivamente recuperato dal Web è estremamente variabile, proprio per la variabilità del livello di conoscenza condivisa delle entità oggetto della richiesta.

Anche se si assiste ad una emergente integrazione e cooperazione su Internet, la mancanza di una condivisione semantica su concetti ed entità rende difficoltoso il reperimento dell'informazione attesa e, qualora questa si riesca a reperire, obbliga gli utenti ad un raffinamento manuale delle informazioni recuperate. Nell'ambito dell'Informazione Geografica la difficoltà è amplificata dalle diverse relazioni (principalmente tassonomiche e topologiche) che caratterizzano la descrizione di features spaziali anche molto simili tra loro, spesso espresse in modo vago e non formalizzate in maniera rigorosa.

In contesti come ad esempio quello delle Spatial Data Infrastructures (SDI), che per definizione forniscono coordinamento all'interno di una comunità con dimensioni politiche, legislative, istituzionali ed umane rilevanti, l'utilizzo di metodologie per la rappresentazione semantica fornisce supporto per attuare processi automatici di scambio ed elaborazione di Informazione Geografica.

L'esplicitazione della conoscenza (knowledge representation) per l'informazione geografica può avvantaggiarsi, come peraltro avviene per altre categorie di informazione presente su Web, dell'approccio basato sulle ontologie. Esse costituiscono le specifiche formali di concetti e relazioni tra concetti relativi ad un certo dominio, basate su un consenso raggiunto all'interno di una determinata "community".

Questo lavoro presenta una metodologia per lo sviluppo di un'ontologia in ambito GIS e come essa possa essere utilizzata a supporto dell'elaborazione automatica attraverso il paradigma dei servizi Web.

## TECNOLOGIE WEB PER L'INFORMAZIONE GEOGRAFICA

Nell'ambito dell'informazione geografica un notevole passo avanti per l'interoperabilità tra sistemi eterogenei è stato fornito dal Geographic Markup Language (GML), una specifica di un linguaggio della famiglia XML, standardizzato dall'OpenGIS Consortium (OpenGIS, 2004) e dotato di uno schema astratto di classi e namespaces per la definizione di entità geografiche.

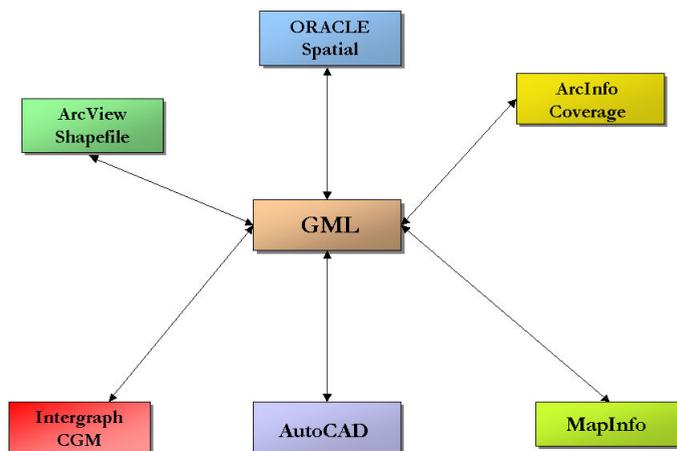


Figure 1: GML e il processo di codifica per differenti formati proprietari di dati spaziali

Allo stesso modo di XML (che consente di separare il contenuto dalla presentazione delle informazioni su Web), il GML fornisce la possibilità di svincolare il dato geografico dallo schema di rappresentazione e dall'algoritmo di generazione delle mappe. Normalmente ciò avviene attraverso lo sviluppo di appositi tool di rendering che possono interpretare i dati geografici esposti in formato GML. Quindi l'approccio è quello di effettuare il rendering delle geometrie descritte dai documenti GML trasformandole in elementi visuali attraverso l'uso di stili XML per la conversione di documenti (tecnologie XSLT). Una specifica per la visualizzazione grafica di dati definita dal W3C, il consorzio per la standardizzazione del Web, è lo Scalable Vector Graphics (SVG). Attraverso un componente di rendering di interpretazione dell'SVG stesso (solitamente un plugin) si ottiene l'immagine visualizzabile su un browser.

La principale modalità di utilizzo del GML supportata dai diversi sistemi GIS è di consentire lo scambio e la trasformazione di dati spaziali di tipo proprietario verso il GML stesso (figura 1), garantendo l'interoperabilità sintattica tra diversi formati di dati spaziali.

Essendo il GML basato su XML non è in grado di portare contenuto semantico per due motivi principali:

- il GML e gli altri linguaggi di tipo XML rappresentano una specifica sintattica, pertanto un documento basato su XML porta con sé un significato implicito nei nomi dei marcatori (tags) espressi in linguaggio naturale (Cover, 1998). Il valore semantico associato al documento non è definito in modo formale ma viene attribuito dal fruitore in base ai propri schemi mentali. Invece per un agente software tutti i tags sono ugualmente non significativi;
- il GML non è componibile con altri linguaggi XML. Le entità descritte con altri linguaggi di tipo XML non possono essere utilizzate all'interno di schemi GML.

Quindi nasce l'esigenza, per i dati codificati in GML, di arricchirli di una descrizione e di regole di interpretazione univoche, comprensibili e processabili da una macchina. Il Semantic Web proposto da Tim Berners Lee (Berners-Lee et al., 2001) si propone di estendere il Web con formalizzazioni e linguaggi per esplicitare la semantica dell'informazione, secondo lo schema di un'architettura a strati (layers) costruita per passi successivi. Il Web diviene quindi un framework logico in cui le informazioni e i concetti sottostanti, definiti in maniera non ambigua sono utilizzabili per usare informazione in maniera automatica (machine to machine) e produrne di nuova tramite inferenza.

La formalizzazione della conoscenza tramite ontologie è l'aspetto più significativo che il Semantic Web fornisce come metodologia per la modellazione dell'Informazione Geografica.

L'ontologia può essere definita come una specifica formale di una concettualizzazione condivisa e più formalmente la si può esplicitare in termini di una tupla  $\langle C, R, I, A \rangle$  dove  $C$  è un insieme di concetti<sup>1</sup>,  $R$  è un insieme di relazioni,  $I$  è un insieme di istanze e  $A$  è un insieme di assiomi (Bruijn, Polleres, 2004).

Esplicitare la conoscenza su un dominio ben specifico secondo un'ontologia significa in sostanza concettualizzare gli elementi del dominio, le loro proprietà e le relazioni che questi possono avere tra loro, secondo un consenso raggiunto, una condivisione non ambigua da parte di una 'community' di quel dominio.

Nell'ambito dell'Informazione Geografica possiamo pensare ad ontologie che riguardano le features ed ontologie che sono relative alle relazioni spaziali, tra le quali principalmente le relazioni di tipo tassonomico, topologico, mereologico e di distanza. Alcune modellazioni concettuali sono state studiate e proposte in letteratura, come ad esempio la formalizzazione

---

<sup>1</sup> La parola classe è spesso usata in luogo di concetto. Le classi sono una concreta rappresentazione di concetti.

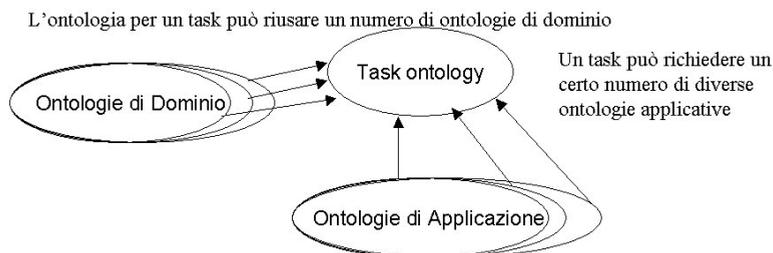
illustrata in (Egenhofer, Herring, 1990) per le relazioni topologiche. Ma la vastità delle problematiche spaziali con i numerosi agganci multidisciplinari lasciano ancora aperte diverse problematiche di modellazione della conoscenza ed ontologica.

## TIPI DI ONTOLOGIE

Per la modellazione di Informazione Geografica, in relazione ad un certo task, ci siamo uniformati alla classificazione illustrata in (Fonseca, Egenhofer, 1999) per le ontologie (figura 2):

- ontologie di dominio: una formalizzazione della conoscenza in un dominio come la topografia, i sistemi di riferimento spaziale, trasporti, reti tecnologiche e così via;
- ontologie di task: una formalizzazione della conoscenza necessaria per la soluzione di un specifico problema a livello più astratto di una specifica situazione o di un contesto all'interno di un'organizzazione (es. l'individuazione di indirizzi);
- ontologie di applicazione: una conoscenza formalizzata per colmare specifiche esigenze applicative non coperte dalle ontologie di dominio.

Questa caratterizzazione ha l'obiettivo di considerare le ontologie come "componenti" di conoscenza riusabili; in tal modo la complessità della progettazione dell'ontologia (che è un'attività necessaria per la formalizzazione di un certo task) può essere ridotta ed affrontata in maniera più sistematica, in analogia alle metodologie di Software Engineering.



**Figura 2:** caratterizzazione delle ontologie

La figura 2 mette in risalto il fatto che diverse ontologie possono essere integrate tra loro per connotare semanticamente un certo task. In particolare l'integrazione e la fusione di diverse ontologie presentano aspetti complessi in quanto comportano la trasformazione tra due o più interpretazioni concettuali e mapping dei termini dall'uno all'altro (Noy, Musen, 2003), nonché aspetti di aggiornamento/modifica nel tempo (versioning).

La progettazione dell'ontologia può partire da schemi standard e rigorosi come quelli tipici dell'UML ma l'espressività dell'ontologia come modellazione concettuale è maggiore proprio in virtù dei gradi di esplicitazione delle relazioni che si possono esprimere tra concetti. Infatti tipicamente dopo aver individuato in modo non ambiguo, termini, concetti e proprietà (solitamente in linguaggio naturale), si modellano aspetti semantici (come le gerarchie, le similarità e la sussunzione<sup>2</sup>) tra concetti.

In particolare la tassonomia risulta essere un caso tipico di ontologia con relazioni di tipo gerarchico (Warren, 2006).

---

<sup>2</sup> Nella logica formale, il sussumere è il ricondurre un concetto nell'ambito di uno più ampio che lo comprende.

L'ultimo passo è rappresentato dalla formalizzazione dell'ontologia attraverso linguaggi specifici. Un linguaggio standard dal W3C per lo sviluppo di ontologie è il linguaggio OWL che permette di caratterizzare classi e proprietà con un esteso numero di operatori (come il costrutto per identificare la similarità tra classi) e di costrutti specifici per definire proprietà tra i concetti. Il linguaggio OWL (OWL, 2004) è basato sulle logiche descrittive<sup>3</sup>, una famiglia di formalismi di rappresentazione della conoscenza di tipo "class-based". Per l'informazione geografica questa maggiore ricchezza descrittiva risulta utile per definire similarità semantiche di relazione tra le feature.

## CASO DI STUDIO

È stato utilizzato un caso abbastanza comune in ambito di applicazioni GIS: l'ottenimento degli indirizzi associati ad un certo codice postale.

Per la progettazione dell'ontologia si fa ricorso ad una specifica standard dell'OpenGIS Consortium sui sistemi di riferimento di coordinate spaziali<sup>4</sup> (formalizzata in linguaggio OWL).

Inoltre è stata prodotta un'ontologia (figura 3) per l'indirizzo di residenza, modellando come proprietà legate al concetto di *Location* i dati relativi a: numero civico, via, zipcode, stato, città. Le due ontologie sono state integrate attraverso la proprietà<sup>5</sup> che un indirizzo abbia una coordinata spaziale secondo un sistema di riferimento.

Il tool utilizzato per la produzione dell'ontologia è il Protégé, uno dei tool attuali più completi per lo sviluppo di applicazioni basate sulle metodologie del Semantic Web (Samson, Ferguson, 2005).

Si tratta di un framework opensource prodotto dalla Stanford University. Inizialmente sviluppato per fornire un editor visuale per la produzione di ontologie, è stato ampliato in funzioni e caratteristiche fino a dare un supporto pieno (attraverso *plugins*) a diversi aspetti del Semantic Web (come ad esempio le logiche descrittive, il controllo della consistenza<sup>6</sup>, il supporto di OWL, l'integrazione di due o più ontologie).

Nel nostro caso di studio sono state definite alcune istanze usando le specifiche del sistema Gauss Boaga con il datum Rome40 e l'European Datum 1950. Abbiamo definito una proprietà *appliedTo* il cui dominio è la classe *SC\_CoordinateReferenceSystem* della ontologia dell'OpenGIS e il codominio è la classe *Location*.

Dall'altra parte la classe *Location* è stata dotata di una proprietà (inversa alla *appliedTo*) *Has\_PositionDefinedBy* legata alla classe *SC\_CoordinateReferenceSystem*.

---

<sup>3</sup> Una base di conoscenza basata sulla Logica descrittiva comprende due componenti: una Tbox per la conoscenza intensionale (non muta nel tempo) e una Abox per la conoscenza estensionale specifica per le istanze del dominio (soggetto a cambiamento). OWL, è una logica descrittiva in quanto i costrutti OWL per le proprietà, uguaglianza tra classi, asserzioni corrispondono ad asserzioni Abox e gli assiomi OWL (subclass, subproperty, ecc.) corrispondono alla conoscenza espressa dalla Tbox.

<sup>4</sup> OpenGeospatial Consortium (2004): Abstract Specification Topic 2, Spatial referencing by coordinates, Project Document 04-046r3 2004-08-16.

<sup>5</sup> Una proprietà che esplicita una relazione tra due classi è definita in OWL come object property, ovvero una funzione che ha un dominio (cioè una classe su cui si applica la proprietà) e un range che è la classe legata alla prima.

<sup>6</sup> Con riferimento alla descrizione e alle proprietà di una classe, il tool consente di controllare se la classe possa essere istanziata ovvero avere elementi che la popolano. Una classe è ritenuta inconsistente se per essa non è possibile avere nessuna istanza.

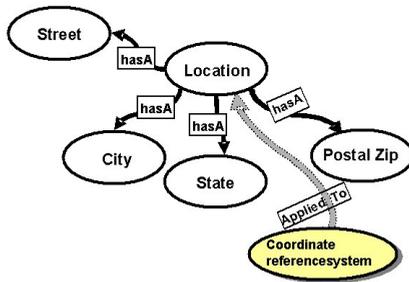


Figura 3: visualizzazione dell'ontologia

La figura 4 mostra un passo della definizione della proprietà di una classe relativamente al nostro caso di studio.

The screenshot shows the Protégé 3.1.1 interface. On the left, the 'CLASS BROWSER' displays a hierarchy of classes under the project 'object1'. The central workspace shows a diagram where 'SpatialThing' is a base class with subclasses 'State', 'StreetAddress', 'Location', 'City', 'Address', and 'PostalZip'. The right pane, 'CLASS EDITOR', is focused on the 'OGC-CS-SC\_CoordinateReferenceSystem' class. The 'AppliedTo' property is being defined, with its value set to 'multiple Location'. The 'Asserted Conditions' pane shows various logical constraints for the class.

Figura 4: Definizione della Object property *appliedTo*

L'ontologia così definita è stata integrata come namespace in un framework per lo sviluppo in ambiente Java di web services semantici chiamato OWL-S IDE (Srinivasan et al., 2005), basato sulla specifica OWL-S (OWL-S, 2004) per la definizione di servizi Web descritti semanticamente

in termini di ontologie. Queste sono utilizzate nell'ambito della specifica OWL-S per coordinare il discovery, la selezione e la composizione di Web Services.

L'ontologia sviluppata è stata utilizzata con successo per soddisfare tramite servizi OWL-S richieste provenienti da un sistema GIS.

In un contesto italiano caratterizzato da processi di e-Government e di reti di interconnessione tra enti pubblici, la metodologia illustrata nel nostro caso di studio può applicarsi alla formulazione di un framework di cooperazione applicativa in ambito GIS mediato dalle tecnologie del Semantic Web. Una possibile strategia nell'ambito di un'applicazione WebGIS potrebbe essere rappresentata dall'aggiunta (a livello di astrazione architeturale) di un *semantic layer* formulato in termini di ontologie di dominio basate su standard tipici dell'informazione geografica (ad esempio FGDC, ISO 19115, 19119 o anche le specifiche dell'IntesaGIS) che possono essere formulate ad esempio in OWL.

In tal modo l'interoperabilità semantica può essere ottenuta più efficacemente attraverso l'elaborazione di informazione mediata da servizi con descrizioni di tipo OWL-S (e OWL-S IDE offre gli strumenti per il discovery e l'interrogazione di servizi semantici) astruendo da sistemi GIS proprietari.

Questo approccio è compatibile anche con la direttiva INSPIRE che in ambito europeo si pone l'obiettivo di stabilire un'infrastruttura per l'Informazione Geografica e di fornire regole per l'accesso, il più aperto possibile, ai dati spaziali di pertinenza di organismi pubblici. D'altra parte l'approccio proposto è allineato anche all'iniziativa definita Geospatial Semantic Web Interoperability Experiment (GSWIE) lanciata dall'OGC nell'aprile 2005 (Lieberman et al., 2005) allo scopo di indirizzare tutta la comunità GIS verso lo sviluppo di un Geospatial Semantic Web.

## CONCLUSIONI

L'effettiva fruibilità di informazione geografica su Web è ostacolata dal fatto che i dati spaziali sono spesso incompatibili tra loro perché creati con propositi e schemi diversi da organizzazioni eterogenee. L'incompatibilità dei diversi sistemi GIS, le difficoltà di condivisione e riuso dei dati geografici rendono difficile l'individuazione dei dati necessari ad un certo task dell'utente e al loro utilizzo integrato.

Molti dei problemi di interoperabilità sono stati affrontati con le tecnologie XML e dei Web Services standard che rendono possibile l'integrazione e l'utilizzo dei diversi sistemi GIS attraverso l'interfaccia del Web. Tale approccio si è rivelato inadeguato poiché non fornisce piena certezza sulla reale condivisione di termini e concetti effettivamente intesi e la semantica dei servizi può essere ambigua<sup>7</sup> nei dati e nei servizi invocati: questo potrebbe avvenire in particolare nel caso di composizione di due o più servizi formalmente compatibili ma semanticamente riferiti a situazioni diverse.

La sperimentazione effettuata mostra le potenzialità delle ontologie e dei linguaggi del Semantic Web per la specifica formale di semantica per l'informazione geografica. Questo tipo di approccio è stato indicato anche recentemente come uno degli elementi essenziali alla costruzione di un Web Semantico (Berners-Lee, 2006).

L'utilizzo di ontologie per la definizione e l'implementazione semantica di Servizi Web può essere di notevole supporto in ambito GIS in tutti quei task che richiedano la composizione di servizi GIS diversi. Si è constatato che la specifica OWL-S può essere utilizzata per modellare il

---

<sup>7</sup> I nomi usati nei file WSDL, usati per la descrizione dei servizi, possono non rappresentare l'esatto significato dei Web Services (per esempio possono essere usati nomi non corrispondenti ai concetti ma nomi prettamente tecnici) e la semantica del Web Service non è realmente descritta nei file WSDL.

servizio come un processo e fornire quindi supporto per la formulazione semantica della composizione di servizi per l'Informazione Geografica.

## **BIBLIOGRAFIA**

Berners-Lee T., Hendler J., Lassila O., 2001. "The Semantic Web", *Scientific American*, May 2001

Berners-Lee T., 2006. "Spread the word, and join it up" in *The Guardian Technology* Aprile 2006, <http://technology.guardian.co.uk/weekly/story/0,,1747327,00.html>

Bruijn de J., Polleres A., 2004. "Towards an Ontology Mapping Specification Language for the Semantic Web", *DERI Technical Report 2004-06-30*, June 2004

Cover R., 1998. "XML and Semantic Transparency", <http://xml.coverpages.org/xmlAndSemantics.html>

Egenhofer M., Herring J., 1990. "A Mathematical Framework for the Definition of Topological Relationships", *Fourth International Symposium on Spatial Data Handling*, K. Brassel and H. Kishimoto (eds) Zurich, Switzerland, pp. 803-813

Fonseca F., Egenhofer M., 1999. "Knowledge Sharing in GIS" in Scheuerman, P. (Ed) *The 3<sup>rd</sup> IEEE International Knowledge and Data Engineering Exchange Workshop*, Chicago, IL, 85-90

Lieberman J., Pehle T., Dean M., 2005. "Semantic Evolution of Geospatial Web Services", [http://www.w3.org/2005/04/FSWS/Submissions/48/GSWS\\_Position\\_Paper.html](http://www.w3.org/2005/04/FSWS/Submissions/48/GSWS_Position_Paper.html)

Noy N. F., Musen M.A. 2003. "The PROMPT Suite: Interactive Tools For Ontology Merging and Mapping", *International Journal of Human-Computer Studies*, 59/6: 983-1024

OpenGIS, 2004. "Geography Markup Language (GML) OGC Document N. 005r3 Version: 3.1.0", [http://portal.opengeospatial.org/files/?artifact\\_id=4700](http://portal.opengeospatial.org/files/?artifact_id=4700)

OWL, 2004. <http://www.w3.org/2004/OWL/>

OWL-S, 2004. "Semantic Markup for Web Services", <http://www.daml.org/services/owl-s/1.1/overview/>

Samson T., Ferguson R., 2005. "Application development with Protégé", *8th Intl. Protégé Conference*, July 18-21, 2005 - Madrid, Spain

Srinivasan N., Paolucci M., Sycara K., 2005. "CODE: A Development Environment for OWL-S Web services", *Tech. Report CMU-RI-TR-05-48*, Robotics Institute, Carnegie Mellon University, October 2005

Warren P., 2006. "Knowledge Management and the Semantic Web: From Scenario to Technology", *IEEE Intelligent Systems*, Vol. 21, Issue 1, Jan. Feb. 2006, 53-59